



HAL
open science

Modélisation et prédiction dynamiques du risque de complications dans le cadre de pathologies hépatiques

Clémence Moreau

► **To cite this version:**

Clémence Moreau. Modélisation et prédiction dynamiques du risque de complications dans le cadre de pathologies hépatiques. Statistiques [math.ST]. Université d'Angers, 2022. Français. NNT : 2022ANGE0031 . tel-04368908

HAL Id: tel-04368908

<https://theses.hal.science/tel-04368908v1>

Submitted on 2 Jan 2024

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.

THÈSE DE DOCTORAT DE

L'UNIVERSITÉ D'ANGERS

ÉCOLE DOCTORALE N° 605

Biologie, Santé

Spécialité : *Mathématiques et leurs interactions*

Par

Clémence MOREAU

Modélisation et prédiction dynamiques du risque de complications dans le cadre de pathologies hépatiques

Thèse présentée et soutenue à Angers, le 20/10/2022

Unité de recherche : EA 3859 - HIFIH Hémodynamique, Interaction Fibrose et Invasivité tumorale Hépatiques

Rapporteurs avant soutenance :

Edouard BARDOU-JACQUET	Professeur d'Université - Praticien Hospitalier, Université de Rennes I
Hélène JACQMIN-GADDA	Directrice de recherche, Université de Bordeaux

Composition du Jury :

Président :	Patrick SAULNIER	Professeur d'Université - Praticien Hospitalier, Université d'Angers
Examinatrice :	Laure ELKRIEF	Maître de Conférences des Universités - Praticienne Hospitalière, Université de Tours
Dir. de thèse :	Jérôme BOURSIER	Professeur d'Université - Praticien Hospitalier, Université d'Angers
Co-dir. de thèse :	Jérémie RIOU	Maître de Conférences des Universités - Praticien Hospitalier, Université d'Angers
Encadrant de thèse :	Marine ROUX	Ingénieure de recherche, Université d'Angers

REMERCIEMENTS

Mes premiers remerciements sont adressés Jérôme Boursier et Jérémie Riou, respectivement directeur et co-directeur de cette thèse. Merci à tous les deux de m'avoir fait confiance il y a maintenant trois ans lors de mon stage de M2, et de m'avoir proposé ce projet. Votre pédagogie ainsi que votre expérience m'auront beaucoup apporté durant cette thèse. Merci Jérôme de m'avoir appris tant de choses à propos de la recherche clinique et de l'hépatologie, malgré mon profil mathématique initial. Jérémie, je te remercie pour ton expertise statistique et tes encouragements dans les moments de doute tout au long de ces années. Je dois une grande partie de mes connaissances actuelles à votre duo complémentaire.

Merci à ma co-encadrante, Marine Roux, sans qui cette thèse n'aurait sûrement pas été la même. Merci pour tes nombreuses relectures qui m'ont aussi permis de progresser. Je sais que ton rêve était d'avoir des remerciements en alexandrins, mais promis je me rattraperai ! Merci pour tes nombreux conseils, et ton aide précieuse tout au long de ces années. Merci d'avoir été une super coloc' de bureau et de m'avoir supporté pendant tous ces jours passés à HIFIH. Merci aussi pour tout ce que tu as fait pour moi (au bureau ou en dehors) et tous les moments passés ensemble. J'espère sincèrement qu'il y en aura d'autres !

Je voudrais ensuite remercier toutes les personnes qui me font l'honneur de faire partie du jury de cette thèse. Merci à Hélène Jacqmin-Gadda et à Edouard Bardou-Jacquet d'avoir accepté de juger ce travail en étant rapporteurs. Merci à Patrick Saulnier d'avoir immédiatement accepté de présider ce jury. Et enfin, merci à Laure Elkrief de m'avoir suivi lors de ces trois années et d'avoir accepté de faire partie de mon jury. Soyez tous assurés de ma profonde gratitude.

Merci à tous les membres de l'équipe HIFIH que j'ai pu croiser et avec qui j'ai pu travailler durant ces travaux de thèse. Une petite mention spéciale à Julien, et nos discussions à propos de la pluie et du beau temps, les matins à l'heure où les bureaux sont encore bien vides.

Certains professeurs vous marquent plus que d'autres : merci à Madame Deloche et Monsieur Hargous de m'avoir transmis le goût pour les Sciences de la Vie et de la Terre et les Mathématiques durant mes années lycée.

Je voudrais ensuite adresser un énorme merci à tous mes amis qui ont été à mes côtés durant ces trois années de thèse. Merci pour votre présence et les grands moments de rigolade. Merci aux nageurs et membres de l'ASA natation (sans oublier le super coach) pour m'avoir accueilli à bras ouverts lorsque je débarquais fraîchement à Angers il y a maintenant cinq années. Certains d'entre vous sont même de véritables amis en dehors des bassins aujourd'hui. Merci pour votre bonne humeur constante et vos encouragements durant la dernière ligne droite de cette thèse. Merci pour tous ces moments passés ensemble, et pour tous les autres à venir.

Et enfin, mes derniers remerciements s'adressent à ma famille. Merci à mon petit frère, alias « Mimile », qui n'est plus si petit que ça maintenant. Merci à mes parents sans qui rien de tout cela n'aurait été réalisable. Merci pour tout ce que vous m'avez apporté depuis toujours, votre soutien et vos encouragements.

SOMMAIRE

Remerciements	iii
Sommaire	v
Liste des Figures	ix
Liste des Tables	ix
Liste des Abréviations	xi
Préambule	1
1 Contexte clinique	5
1.1 La <i>Non-Alcoholic Fatty Liver Disease</i>	5
1.2 L'hépatite C	9
1.3 Évaluation de la fibrose hépatique	12
1.3.1 La biopsie	12
1.3.2 Les tests non-invasifs	14
1.3.2.1 Tests sanguins	14
1.3.2.2 Élasticité hépatique	15
1.4 Cohortes	17
2 État de l'art statistique	19
2.1 Modèle linéaire généralisé à effets mixtes	19
2.1.1 Modèle linéaire à effets mixtes	20
2.1.2 Modèle linéaire généralisé à effets mixtes (GLMM)	23
2.1.2.1 Définition du modèle linéaire généralisé à effets mixtes	23
2.1.2.2 Estimation des paramètres	23
2.1.3 Modèle mixte à classes latentes	25
2.1.3.1 Définition du modèle mixte à classes latentes	25
2.1.3.2 Estimation des paramètres	26

2.2	Analyse de survie : généralités	27
2.2.1	Fonctions associées aux distributions de survie	27
2.2.2	Troncature et censures	29
2.2.2.1	Troncature à gauche	29
2.2.2.2	Censure à droite	30
2.2.2.3	Censure par intervalle	31
2.2.3	Courbe de survie	31
2.2.3.1	Estimateur de Kaplan-Meier	31
2.2.3.2	Test du Logrank	33
2.3	Modèle à risques proportionnels : le modèle de Cox	35
2.3.1	Modèle de Cox	35
2.3.1.1	Définition du modèle	35
2.3.1.2	Estimation du paramètre	36
2.3.1.3	Interprétation des coefficients du paramètre	37
2.3.2	Modèle de Cox dépendant du temps	38
2.3.3	Modélisation landmark : un raffinement du modèle de Cox	39
2.4	Risques concurrents	40
2.4.1	Fonctions associées aux distributions de survie	41
2.4.1.1	Définitions	41
2.4.1.2	Estimations	42
2.4.2	Modèles de régression	43
2.4.2.1	Régression sur des risques cause-spécifique	43
2.4.2.2	Régression sur plusieurs types d'évènements	44
2.4.2.3	Régression sur les fonctions d'incidence cumulée	44
2.5	Modèle conjoint	46
2.5.1	Modèle conjoint à effets aléatoires partagés	47
2.5.1.1	Définition du modèle conjoint à effets aléatoires partagés	47
2.5.1.2	Estimation des paramètres	48
2.5.2	Modèle conjoint à classes latentes	49
2.5.2.1	Définition du modèle conjoint à classes latentes	49
2.5.2.2	Estimation des paramètres	50
2.5.2.3	<i>Expected Prognostic Observed Cross-Entropy</i>	51
2.6	Critères de performance	52
2.6.1	Évaluation des capacités de discrimination d'un marqueur	53

2.6.1.1	Courbe ROC et AUC	53
2.6.1.2	C-index	56
2.6.2	Évaluation des capacités de calibration d'un modèle : le Brier Score	56
2.6.3	Évaluation de prédictions dynamiques	57
2.6.3.1	Prédictions dynamiques	58
2.6.3.2	Critères de performance dynamiques	61
2.7	Multiplicité des tests	62
2.7.1	Taux d'erreur	63
2.7.1.1	Erreur de type I	63
2.7.1.2	Erreur de type II	64
2.7.2	Correction de la multiplicité	65
3	Stratification du risque de complications hépatiques par des tests non-invasifs chez les patients atteints de NAFLD	67
3.1	Article scientifique	67
3.2	Documents complémentaires	77
4	Prédiction personnalisée du risque de complications hépatiques après une RVS chez les patients infectés par le VHC	91
4.1	Article scientifique	91
4.2	Documents complémentaires	117
5	Comparaison de plusieurs critères de performance pour prédictions dynamiques	131
5.1	Article scientifique	132
5.2	Documents complémentaires	149
6	Discussion	157
6.1	Résumés et perspectives des travaux de thèse	157
6.2	Nouvelle approche pronostique	159
	Bibliographie	165
	Valorisation scientifique et enseignements	175

LISTE DES FIGURES

1	De la prise en charge statique vers la prise en charge dynamique	2
1.1	Nombre de publications associées au mot-clé « NAFLD » référencées dans PubMed	6
1.2	Histoire naturelle de la NAFLD	7
1.3	Répartition géographique de la NAFLD (YOUNOSSI et al. 2019, Fig. 1) . .	8
1.4	Nombre de publications associées au mot-clé « Hepatitis C » référencées dans PubMed	10
1.5	Histoire naturelle de l'infection par le VHC	10
1.6	Échantillons de biopsie hépatique observés à différents stades de fibrose selon le score METAVIR : F1 (A), F2 (B), F3 (C) et F4 (D)	13
1.7	Mesure de l'élasticité hépatique à l'aide du Fibroscan	16
2.1	À gauche : modèle sans effet aléatoire ; au centre : modèle avec un intercept aléatoire ; à droite : modèle avec une pente aléatoire	21
2.2	À gauche : matrice de variance-covariance non-structurée ; à droite : matrice de variance-covariance diagonale	21
2.3	Suivi de cinq patients entre 2004 et 2016	29
2.4	Courbes de survie obtenues à l'aide de l'estimateur de Kaplan-Meier	33
2.5	Modélisation landmark	39
2.6	Diagramme d'un modèle à K risques concurrents, $k \in \{1, \dots, K\}$, $K \geq 2$.	40
2.7	Courbes d'incidence cumulée pour deux évènements concurrents observés .	43
2.8	Graphique de la structure d'un modèle conjoint (FERRER 2017, Fig. 2.4) .	46
2.9	Courbe ROC (A) et AUCs dépendants du temps (B)	55
2.10	Prédictions dynamiques calculées aux temps landmark $s = 2$ (A) et $s = 6$ ans (B) pour des horizons de $h = \{1, 2, 3, 4\}$ ans	60
6.1	Changement de paradigme dans la gestion des maladies hépatiques chroniques	161

LISTE DES TABLES

1.1	Définition du syndrome métabolique (ALBERTI, ZIMMET et SHAW 2005)	6
1.2	<i>NASH Clinical Research Network Score</i> (KLEINER et al. 2005)	13
1.3	Tests sanguins simples et spécialisés utilisés dans le cadre de la détection de fibrose hépatique	15
2.1	Table de contingence du nombre d'évènements et de l'effectif à risque par groupe au temps t_j	34

LISTE DES ABRÉVIATIONS

AASLD	<i>American Association for the Study of Liver Diseases</i>
ALT	ALanine aminoTransf�rase
ANRS	Agence Nationale de Recherches sur le Sida et les h�patites virales
APRI	<i>AST Platelet Ratio Index</i>
ARN	Acide RiboNucl�ique
AST	ASpartate aminoTransf�rase
AUC	<i>Area Under the ROC Curve</i>
BIC	<i>Bayesian Information Criterion</i>
CHC	Carcinome H�patoCellulaire
CVPOL	<i>Cross-Validated Prognosis Observed Loss</i>
EASL	<i>European Association for the Study of the Liver</i>
ELF	<i>Enhanced Liver Fibrosis</i>
eLIFT	<i>easy LIver Fibrosis Test</i>
EPOCE	<i>Expected Pronostic Observed Cross-Entropy</i>
FDR	<i>False Discovery Rate</i>
FIB-4	<i>FIBrosis-4</i>
FWER	<i>Family-Wise Error Rate</i>
GGT	Gamma GlutamylTransf�rase

HR	<i>Hazard Ratio</i>
i.i.d.	indépendants et identiquement distribués
IMC	Indice de Masse Corporelle
IPA	<i>Index of Prediction Accuracy</i>
IPCW	<i>Inverse Probability of Censoring Weighting</i>
IQR	<i>InterQuartile Range</i>
kPa	kiloPascal
MPOL	<i>Mean Prognosis Observed Loss</i>
NAFL	<i>Non-Alcoholic Fatty Liver</i>
NAFLD	<i>Non-Alcoholic Fatty Liver Disease</i>
NAS	<i>NAFLD Activity Score</i>
NASH	<i>Non-Alcoholic SteatoHepatitis</i>
NFS	<i>NAFLD Fibrosis Score</i>
OMS	Organisation Mondiale de la Santé
PIIIP	<i>amino-terminal Propeptide of type III Procollagen</i>
ROC	<i>Receiver Operating Characteristic</i>
RVS	Réponse Virologique Soutenue
TIMP-1	<i>Tissue Inhibitor of matrix MetalloProteinase 1</i>
TP	Taux de Prothrombine
VCTE	<i>Vibration Controlled Transient Elastography</i>
VHC	Virus de l'Hépatite C
VIH	Virus de l'Immunodéficience Humaine

PRÉAMBULE

La prise en charge des patients ayant une maladie chronique du foie nécessite d'évaluer la sévérité des lésions hépatiques. Parmi toutes celles observées, c'est le degré de fibrose du foie qui est le plus corrélé au pronostic du patient. L'examen de référence pour l'évaluation de ces lésions est la biopsie hépatique. Cependant, c'est une méthode invasive non sans risque, avec des biais d'échantillonnage et d'observateur. Aujourd'hui, les méthodes de diagnostic non-invasif hépatique sont préférées. Elles connaissent néanmoins des performances diagnostiques limitées. Un moyen de palier à ces difficultés serait d'évoluer vers un changement de paradigme dans la prise en charge du patient : ne plus agir en fonction de l'état lésionnel du foie mais plutôt selon le pronostic du patient. Ainsi, comprendre la dynamique des maladies hépatiques est un enjeu essentiel. Cette thèse s'inscrit dans cette démarche résumée en Figure 1. Nous nous intéressons dans ce manuscrit à deux types de maladies chroniques du foie : la *Non-Alcoholic Fatty Liver Disease* (NAFLD) et l'hépatite C chronique. Dans le cadre de la NAFLD, nous cherchons à évaluer l'histoire naturelle de maladie où le risque de complications hépatiques est croissant avec le temps. Pour l'hépatite C chronique, nous nous plaçons après l'éradication du virus par un traitement antiviral et étudions le risque résiduel de faire une complication.

L'objectif principal de cette thèse est de mettre en œuvre une modélisation dynamique du risque de complications liées au foie dans le cadre de pathologies hépatiques. Une telle modélisation permet ainsi de prendre en compte tout le suivi du patient afin d'ajuster et de mettre à jour les prédictions de risque à un horizon fixé. La biopsie est la référence pour mesurer le stade de fibrose hépatique mais compte tenu des risques qu'elle comporte, cet examen ne peut être répété souvent dans le temps. Moins coûteux et sans risque pour le patient, les tests non-invasifs hépatiques peuvent être répétés régulièrement durant le suivi à l'aide d'une simple prise de sang ou d'un examen d'élastométrie. Ainsi, les prédictions de risque obtenues à l'aide de modèles dynamiques peuvent être actualisées à chaque nouvelle visite du patient.

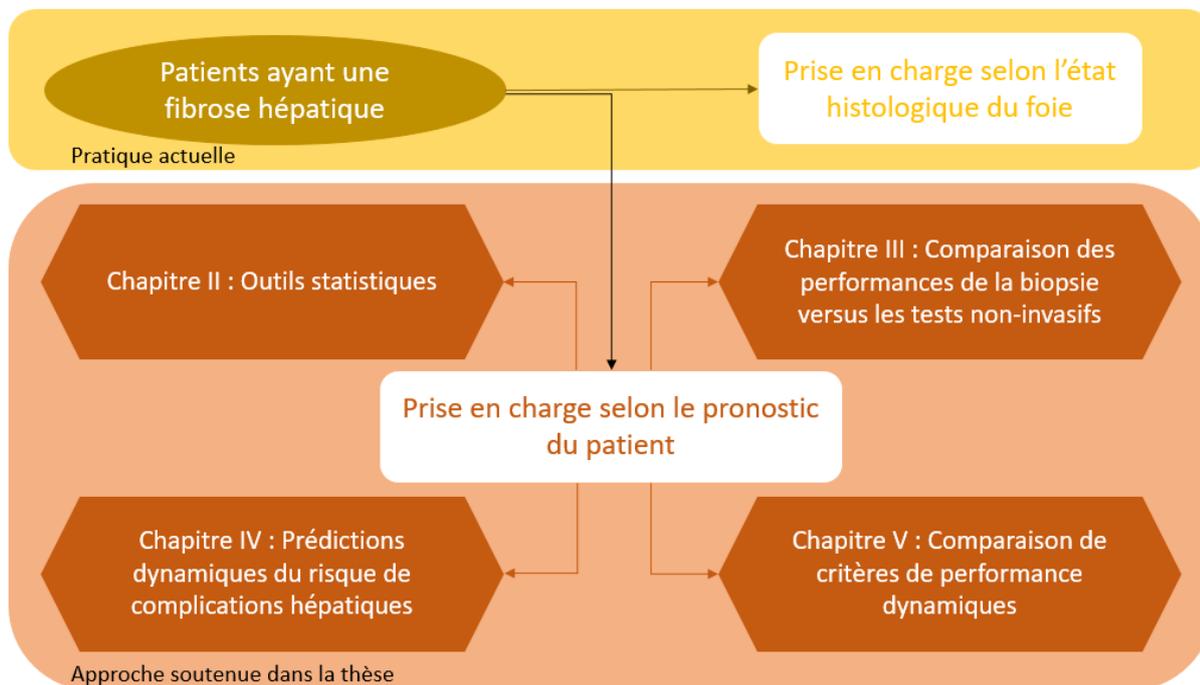


FIGURE 1 – De la prise en charge statique vers la prise en charge dynamique

En premier lieu, le Chapitre 1 présente le contexte clinique de cette thèse. Il décrit les différentes caractéristiques de la NAFLD et de l'hépatite C chronique (définition, histoire naturelle, données épidémiologiques, facteurs de risque et traitements), ainsi que les différentes méthodes de détection de la fibrose hépatique.

Par la suite, le Chapitre 2 présente une revue de la littérature des différents outils statistiques abordés au cours de cette thèse. Il permet d'introduire la théorie des modèles à effets mixtes, des modèles de survie, puis des modèles conjoints. Ce chapitre expose également différents critères de performance permettant d'évaluer la qualité de discrimination et de calibration de différents modèles statistiques ou biomarqueurs. Enfin, cette partie propose une rapide introduction à la multiplicité des tests.

Les Chapitres 3, 4 et 5 sont présentés sous la forme d'articles scientifiques : certains sont en cours de publication, d'autres publiés. Le Chapitre 3 compare les capacités pronostiques de la biopsie du foie face à celles des tests non-invasifs chez des patients atteints de NAFLD. Il montre aussi la bonne stratification du risque de complications hépatiques par un algorithme combinant un test sanguin et l'élastométrie chez ces malades.

Ces tests non-invasifs permettent d’avoir un suivi régulier de la fibrose : ils sont utilisés dans le Chapitre 4 afin de développer un modèle dynamique pour prédire les complications hépatiques chez des patients guéris du virus de l’hépatite C. Ce modèle a été conçu à l’aide des cohortes de dérivation ANRS CO12 CirVir et de validation ANRS CO22 HEPATHER, décrites dans la Section 1.4.

Plus méthodologique, le Chapitre 5 introduit un nouvel outil statistique permettant la comparaison de plusieurs critères de performance dynamiques afin de déterminer le meilleur biomarqueur pronostique à suivre dans le temps ou le meilleur modèle dynamique. Avec les progrès technologiques de ces dernières années en pratique clinique, de plus en plus de biomarqueurs peuvent être mesurés (le Fibroscan en hépatologie par exemple). Nous avons élaboré une stratégie permettant de choisir lesquels sont les plus pertinents à suivre dans le temps afin de prédire un évènement d’intérêt. Nous avons ensuite appliqué la méthode chez des patients atteints de NAFLD, en comparant les capacités pronostiques de trois tests sanguins et du Fibroscan pour prédire des complications hépatiques.

Enfin, le Chapitre 6 résume les différents résultats obtenus lors de ces trois années de thèse. Il propose également différentes perspectives envisagées suite aux travaux effectués, ainsi qu’une nouvelle approche pronostique pour les patients atteints de maladies hépatiques chroniques.

CONTEXTE CLINIQUE

Dans un premier temps, ce chapitre présente les caractéristiques et l'histoire naturelle des deux pathologies hépatiques étudiées durant cette thèse : la NAFLD et l'hépatite C chronique. Les données épidémiologiques, les facteurs de risque et les traitements associés à chacune d'entre elles y sont détaillés. Par la suite, ce même chapitre propose une revue des différentes méthodes utilisées en pratique afin de détecter la fibrose hépatique. Enfin, les cohortes ayant servi pour répondre aux objectifs énoncés en préambule de cette thèse sont également décrites.

1.1 La *Non-Alcoholic Fatty Liver Disease*

Aussi appelée « la maladie du foie gras » par le grand public, la *Non-Alcoholic Fatty Liver Disease* (NAFLD) est une maladie qui touche principalement les patients en situation d'obésité ou ayant un régime alimentaire trop gras (LUDWIG et al. 1980). Elle est présentée comme l'atteinte hépatique de la maladie métabolique (Table 1.1) liée à l'obésité et l'insulinorésistance (ALBERTI, ZIMMET et SHAW 2005). La NAFLD atteint 25% de la population mondiale et représente désormais la première cause des pathologies hépatiques chroniques (YOUNOSSI et al. 2011). L'intérêt pour cette pathologie est exponentiel depuis le début des années 2000 : plus de 5000 publications en rapport avec la NAFLD ont été enregistrées en 2021 dans PubMed, contre environ 500 en 2009 (Figure 1.1). Le fardeau de cette maladie et l'arrivée de nouveaux traitements à l'étude en Phase 2 et 3 contribuent tous les deux au fort intérêt de la communauté scientifique pour cette pathologie.

La NAFLD survient lorsqu'une quantité trop importante de graisse liée à l'insulinorésistance est stockée au niveau du foie. Elle est définie par la présence de vésicules de lipides dans au moins 5% des hépatocytes (KLEINER et al. 2005). Le diagnostic de cette maladie nécessite la mise en évidence de cette stéatose par imagerie ou par histologie (SCHWENZER et al. 2009; RATZIU et al. 2005). En aucun cas elle ne doit être

TABLE 1.1 – Définition du syndrome métabolique (ALBERTI, ZIMMET et SHAW 2005)

Syndrome métabolique (présence d'au moins 3 critères sur les 5)	
Tour de taille	≥ 94 cm pour les hommes ≥ 80 cm pour les femmes
Tension artérielle	≥ 130 mm Hg en systolique ou ≥ 85 mm Hg en diastolique ou traitement antihypertenseur
Triglycérides	≥ 1.5 g/L ou ≥ 1.7 mmol/L ou traitement hypolipémiant
Cholestérol HDL	< 0.4 g/L ou < 1.0 mmol/L pour les hommes < 0.5 g/L ou < 1.3 mmol/L pour les femmes ou traitement hypolipémiant
Glycémie	≥ 1.0 g/L ou ≥ 5.6 mmol/L ou traitement antidiabétique

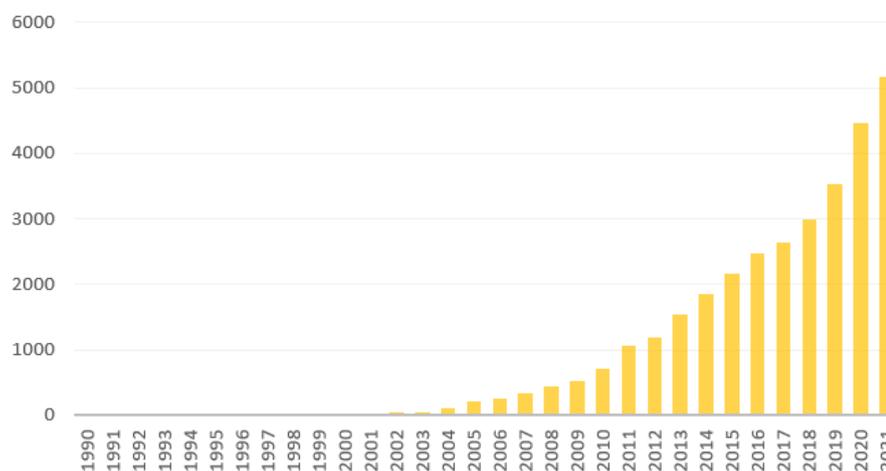


FIGURE 1.1 – Nombre de publications associées au mot-clé « NAFLD » référencées dans PubMed

provoquée par des traitements stéatogènes (corticostéroïdes, méthotrexate, amiodarone ou tamoxifène), une consommation importante d'alcool, ou autres pathologies hépatiques chroniques (CHALASANI et al. 2018).

La NAFLD est susceptible d'évoluer et de s'aggraver au fil du temps avec une progression généralement lente (ANGULO 2002; CLARK, BRANCATI et DIEHL 2002). L'histoire naturelle de la NAFLD est présentée en Figure 1.2. Le stade le plus bénin est celui de la stéatose hépatique non alcoolique (*Non-Alcoholic Fatty Liver* - NAFL). À ce stade, la stéatose est isolée ou accompagnée d'une inflammation lobulaire minime. La pathologie peut ensuite progresser vers la stéatohépatite non alcoolique (*Non-Alcoholic SteatoHepatitis* - NASH). Dans ce cas, la stéatose est accompagnée d'une inflammation lobulaire et d'une ballonnisation des hépatocytes. La NASH favorise l'activation de cellules stellaires qui produisent alors de la fibrose qui s'accumule dans le parenchyme hépatique jusqu'au stade de fibrose avancée, puis à la cirrhose. À ce stade ultime, plusieurs complications peuvent survenir telles que l'hémorragie digestive par rupture de varice, l'ascite, l'insuffisance hépatique, l'insuffisance rénale ou encore la cancer primitif du foie (Carcinome Hépatocellulaire - CHC) (BYRNE et TARGHER 2015).



FIGURE 1.2 – Histoire naturelle de la NAFLD

La Figure 1.3 montre les différents pourcentages de population touchée selon les différentes régions dans le monde (YOUNOSSI et al. 2019, Fig. 1). Alors qu'elle ne touche qu'un peu plus de 13% de la population africaine, la NAFLD atteint plus de 31% de la population du Moyen-Orient et 30% pour l'Amérique du Sud. Presque un quart des populations d'Europe et d'Amérique du Nord vivent avec cette maladie, et plus de 27% pour la population asiatique. Certaines études américaines prévoient une augmentation du nombre de cas NAFLD aux États-Unis de 21% en 2030 par rapport à 2015, touchant des patients de plus en plus jeunes à cause de l'inflation de l'obésité chez les adolescents (ESTES et al. 2018; ALLEN et al. 2018). En résumé, la NAFLD touche aujourd'hui près de 25% de la population mondiale et constitue un réel problème de santé publique à l'échelle planétaire : la prévalence ne cesse d'augmenter et les malades sont de moins en moins âgés.

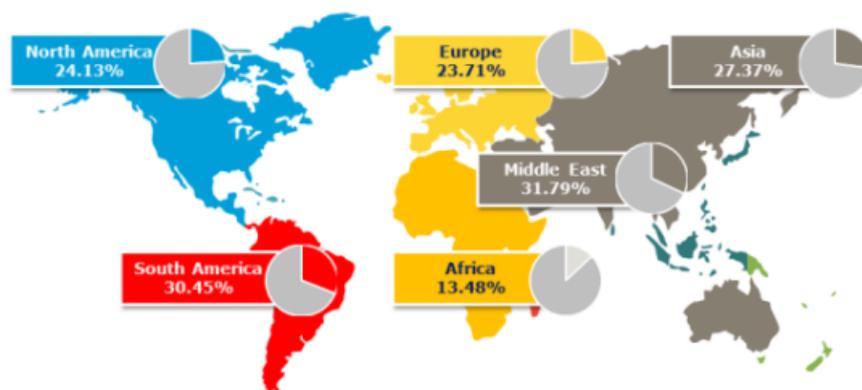


FIGURE 1.3 – Répartition géographique de la NAFLD (YOUNOSSI et al. 2019, Fig. 1)

Les facteurs de risque de la NAFLD sont multiples. Comme cette pathologie est associée à l'obésité et à l'insulinorésistance (BYRNE et TARGHER 2015), la NASH et la sévérité de la fibrose hépatique sont associées à la présence des conditions suivantes : surpoids, diabète, hypertension artérielle, dyslipidémie (cholestérol HDL bas et triglycérides élevées). Il a également été montré que les maladies cardiovasculaires sont elles aussi étroitement liées avec la NAFLD en étant la cause de décès la plus fréquente chez les malades (ANSTEE, TARGHER et DAY 2013). Enfin, le terrain génétique, une mauvaise hygiène de vie (sédentarité, alimentation déséquilibrée), le statut hormonal, l'âge, et le microbiote intestinal vont favoriser l'obésité et l'insulinorésistance qui vont eux-mêmes entraîner la NASH (YOUNOSSI 2019). La revue de COTTER et RINELLA 2020, Fig. 1 détaille les différents facteurs de risque regroupant les comorbidités métaboliques, le microbiome et les facteurs environnementaux, génétiques et épigénétiques.

À ce jour, aucun traitement médicamenteux n'est approuvé contre la NAFLD par les agences de santé mais plusieurs pistes de développement sont en cours (PERAZZO et DUFOUR 2017). Cependant, différentes stratégies existent afin d'améliorer le pronostic des patients, en ralentissant la progression de la maladie et limitant ainsi le risque de complications de cirrhose. Tout d'abord, les patients sont encouragés à modifier leur mode de vie en adoptant un régime alimentaire sain et une activité physique régulière : la perte de poids est le meilleur traitement établi pour la NAFLD (ROMERO-GÓMEZ, ZELBERSAGI et TRENELL 2017). Cela peut améliorer l'état général du foie, les bilans biologiques hépatiques, les taux d'insuline et la qualité de vie des malades. La chirurgie bariatrique peut également être effectuée afin d'obtenir une amélioration progressive et durable de la fibrose hépatique due à la NAFLD (LASSAILLY et al. 2015). Très efficace pour la perte

de poids, elle est recommandée dès 2009 par la Haute Autorité de Santé et permet aussi une amélioration des comorbidités telles que les maladies cardiovasculaires ou le diabète de type 2 (SJÖSTRÖM et al. 2004). Cependant, très peu de patients arrivent à atteindre et maintenir les objectifs de perte de poids avec le régime alimentaire, et la chirurgie bariatrique n'est possible que pour un très faible nombre de patients : cela renforce l'intérêt de la recherche de nouveaux traitements pharmacologiques.

La NAFLD est diagnostiquée lors d'un contexte dysmétabolique mais l'apparition d'une stéatose hépatique peut provenir d'autres causes. En effet, cette stéatose peut aussi être associée à la prise de médicaments (glucocorticoïdes, tamoxifène, amiodarone, méthotrexate) ou à des maladies infectieuses comme le virus de l'hépatite C ou le Virus de l'Immunodéficience Humaine (VIH). La Section 1.2 fait le point sur l'épidémiologie, et les caractéristiques cliniques et pronostiques de l'infection par le virus de l'hépatite C.

1.2 L'hépatite C

L'hépatite C chronique a été reconnue comme un problème majeur de santé publique par l'Organisation Mondiale de la Santé (OMS) au début des années 2000. La prévalence a largement diminué depuis l'introduction de médicaments antiviraux à action directe en 2014 (GOLDBERG et al. 2017). Ces différents phénomènes sont illustrés en Figure 1.4 : le nombre de publications associées au Virus de l'Hépatite C (VHC) dans PubMed a constamment augmenté depuis le début des années 1990 jusqu'en 2014. La grande efficacité de ces nouveaux traitements contre le VHC permet d'obtenir un meilleur pronostic pour les patients infectés. De plus, ils ont permis à l'OMS de mettre en place une stratégie visant à réduire le nombre d'infections par le VHC de 90% et le nombre de décès dus à l'infection de 65% d'ici 2030 (WHO 2021). Pour cela, l'idée est de sensibiliser davantage le public, et d'étendre les services de dépistage, de soins et de traitements pour cette maladie.

La Figure 1.5 présente l'évolution naturelle de l'infection par le VHC (WESTBROOK et DUSHEIKO 2014). Le premier stade, appelé hépatite C aiguë, est celui de l'infection par le VHC. En général, cette phase dure 2 à 12 semaines, les patients n'ont pas de symptôme et la maladie est bénigne. Après avoir été infectés par le VHC, environ 30% des patients ont une guérison spontanée et les 70% restants développent une infection

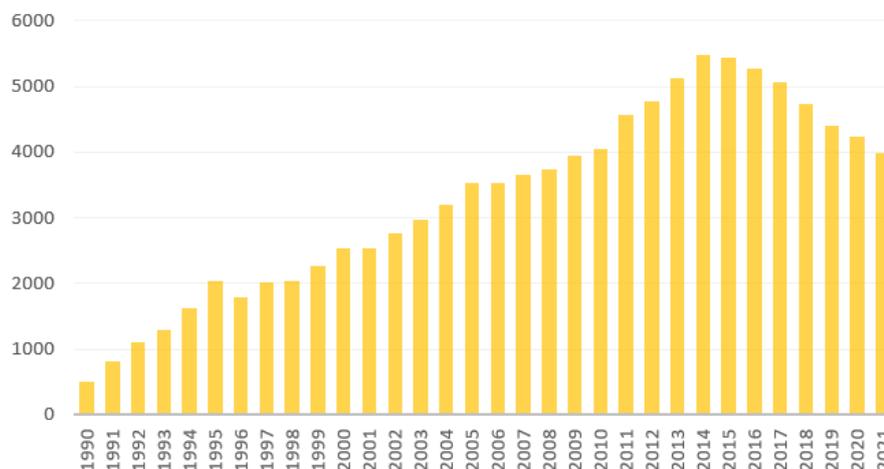


FIGURE 1.4 – Nombre de publications associées au mot-clé « Hepatitis C » référencées dans PubMed

chronique de l'hépatite C. Elle se caractérise par la persistance de l'Acide RiboNucléique (ARN) du VHC dans le sang pendant une durée de 6 mois minimum après le début de l'infection (LINGALA et GHANY 2015). Une fois devenue chronique, l'infection nécessite un traitement et ne peut se résorber seule. À ce stade, la maladie devient permanente et peut entraîner une fibrose progressive allant jusqu'à la cirrhose. Environ 20% des patients infectés chroniquement développent les mêmes complications (ruptures de varice, ascites, insuffisances hépatiques, CHC) que les patients cirrhotiques atteints de la NAFLD sur une période de 20 à 30 ans (BRUDEN et al. 2017).

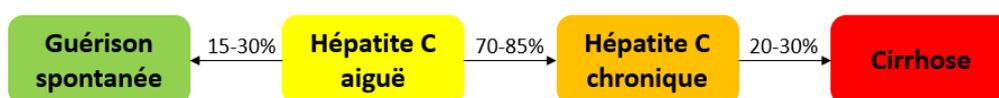


FIGURE 1.5 – Histoire naturelle de l'infection par le VHC

À cause des faibles symptômes pour la majorité des patients infectés, l'hépatite C est peu diagnostiquée : peu de pays disposent d'informations précises sur la prévalence de la maladie à l'échelle nationale. L'Observatoire Polaris a alors été créé en 2015 afin d'estimer la prévalence globale de l'infection par le VHC en recueillant des données issues de 100 pays représentant plus de 85% de la population mondiale (BLACH et al. 2017). Les résultats obtenus ont ensuite fait l'objet d'un rapport mondial sur l'hépatite C en 2017 par l'OMS (WHO 2017). Il a été estimé que 71.1 [95% : 62.1-79.0] millions d'individus

représentant 1.0% [95% : 0.8-1.1] de la population mondiale vivaient avec une hépatite C chronique. C'est une pathologie qui touchent tous les pays : la région la plus fortement touchée est l'Asie du Sud-Est avec 10.4 [95% : 8.0-17.8] millions de malades, tandis que l'Amérique est le continent le plus préservé avec 7.2 [95% : 6.1-8.0] millions de cas. De plus, il a été établi que 29,7% des décès dus à une hépatite en 2015 faisaient suite à une infection par le VHC précédemment.

Le VHC se transmet principalement par le sang lors de transfusions sanguines non-contrôlées, de consommation de drogues injectables, ou d'utilisation de matériel médical non-stérilisé (COOKE et al. 2019). Dans des cas plus rares, le virus peut également se transmettre par voie sexuelle (CHAN et al. 2016) ou bien de la mère à l'enfant (ROBERTS et YEUNG 2002). La progression de l'hépatite C chronique chez les patients infectés est influencée par différents facteurs pouvant être liés aux caractéristiques ou à l'environnement du patient (LINGALA et GHANY 2015, Table 1). Ainsi, l'âge au moment de l'infection, le sexe, l'obésité, la résistance à l'insuline, le diabète de type 2, le génotype, la co-infection par l'hépatite B ou le VIH, la consommation d'alcool ou de drogues sont identifiés comme des facteurs pouvant augmenter le risque de fibrose ou de cirrhose.

Les personnes infectées par le VHC sont diagnostiquées à l'aide d'un test sérologique afin de trouver des traces d'anticorps anti-VHC dans le sang. Les patients sont souvent pris en charge à un stade avancé de la maladie puisque l'infection est asymptomatique pendant des années avant que des complications hépatiques ne surviennent (WESTBROOK et DUSHEIKO 2014). Le but des différents traitements contre le VHC est d'obtenir une Réponse Virologique Soutenue (RVS) : l'ARN du VHC n'est plus détectable chez le patient douze semaines après la fin du traitement (EASL 2018). La RVS est associée à une guérison de l'hépatite C et une amélioration de la fibrose hépatique chez les patients non-cirrhotiques. Les recommandations et les médicaments disponibles sur le traitement de l'hépatite C sont décrits par l'*European Association for the Study of the Liver* (EASL 2018) et l'*American Association for the Study of Liver Diseases* (AASLD 2015).

Ces dernières années, les protocoles de traitement contre l'hépatite C ont connu de grandes transformations grâce à la dernière génération d'antiviraux à action directe (BETHA et al. 2018). Aujourd'hui, plus de 95% des patients traités avec ces nouveaux médicaments contre le VHC obtiennent une RVS et guérissent (SPEARMAN et al. 2019). Cela est alors associé à un risque moindre de décompensation de cirrhose, de CHC ou de décès. Cependant, les recommandations préconisent un suivi des patients cirrhotiques (patients

ayant une fibrose hépatique à un stade avancé), pour lesquels le risque de complications hépatiques subsiste (CARRAT et al. 2019). Néanmoins, la présence de comorbidités chez les patients non-cirrhotiques (consommation d’alcool, terrain dysmétabolique) peut aussi favoriser la poursuite de la progression de la fibrose du foie malgré l’éradication du virus.

1.3 Évaluation de la fibrose hépatique

Que ce soit pour la NAFLD ou pour l’hépatite C chronique, le pronostic des patients est lié au degré de la fibrose hépatique. Il est donc important dans ces deux maladies d’effectuer une évaluation de la fibrose hépatique afin de mettre en place un suivi et des traitements adaptés aux patients. Différentes méthodes existent pour mesurer la fibrose et sont présentées dans cette section.

1.3.1 La biopsie

L’examen de référence pour l’évaluation de la fibrose hépatique chez les patients atteints de NAFLD ou d’hépatite C chronique est la biopsie du foie (CHALASANI et al. 2018). Cependant, il s’agit d’une méthode invasive, coûteuse, non sans risque (1/1000 complication sévère telles que des hémorragies ou perforations d’organe ; 1/10000 décès), et avec des biais d’échantillonnage et d’observateur entraînant une variabilité de la mesure et du résultat (CASTÉRA et al. 2005). En effet, la répartition microscopique de la fibrose hépatique est hétérogène dans le foie et varie donc selon le site de prélèvement (DAVISON et al. 2020). Par ailleurs, l’appréciation du degré de fibrose par l’œil de l’anatomopathologiste peut varier avec une reproductibilité inter-observateur suboptimale.

Afin d’évaluer l’existence de lésions fibreuses dans le foie, les anatomopathologistes utilisent des scores décrivant le degré de fibrose hépatique selon plusieurs stades : les scores de Batts-Ludwig (BATTS et LUDWIG 1995), d’Ishak (ISHAK 1995) et METAVIR (THE FRENCH METAVIR COOPERATIVE STUDY GROUP 1994). En France, le score METAVIR est le plus utilisé dans le cadre de l’hépatite C chronique. Il est composé de cinq stades (F0-4) allant de l’absence de fibrose à la cirrhose : pas de fibrose F0, fibrose portale sans septa F1, fibrose portale et quelques septa F2, nombreux septa sans cirrhose F3, et cirrhose F4. La Figure 1.6 présente quatre échantillons de foie correspondant aux quatre stades avec présence de fibrose F1-4. En pratique, on parle de fibrose significative pour les stades F2-4 et de fibrose sévère pour les stades F3-4.

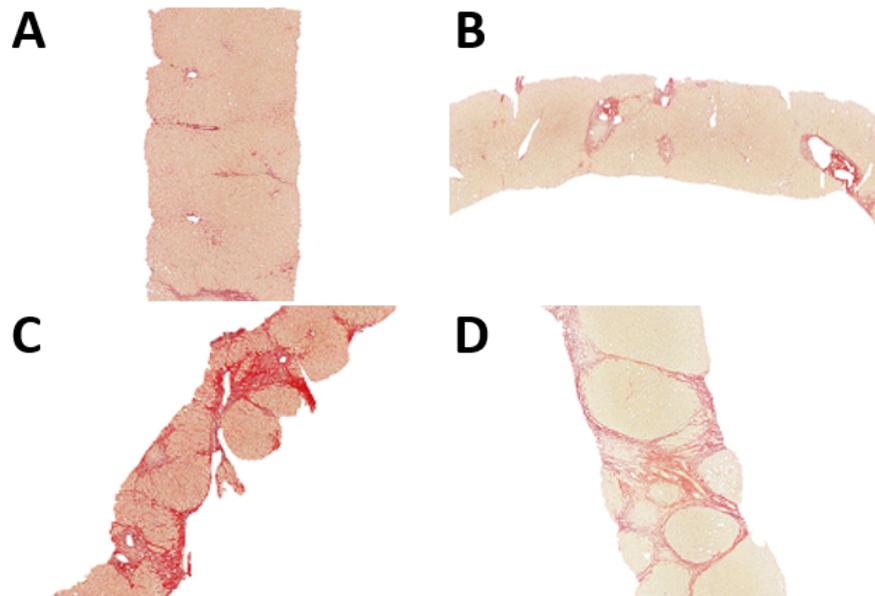


FIGURE 1.6 – Échantillons de biopsie hépatique observés à différents stades de fibrose selon le score METAVIR : F1 (A), F2 (B), F3 (C) et F4 (D)

Un score spécifique pour l'évaluation de la fibrose hépatique dans la NAFLD a été développé par le *NASH Clinical Research Network* (KLEINER et al. 2005). Ce score est détaillé dans la Table 1.2.

TABLE 1.2 – *NASH Clinical Research Network Score* (KLEINER et al. 2005)

Stade	Fibrose
0	Absence
1	Périsinusoïdale ou portale
2	Périsinusoïdale et portale
3	Septale
4	Cirrhose

Même si elle est la méthode de référence pour l'évaluation des lésions hépatiques, il est impossible d'utiliser la biopsie en première ligne pour le diagnostic de fibrose en raison des coûts et des risques liés à l'invasivité de celle-ci. Par exemple, il serait disproportionné de biopsier tous les patients NAFLD représentant un quart de la population mondiale par rapport au nombre de malades qui développent une fibrose avancée (BOURSIER et al. 2016). Aujourd'hui, des méthodes dites non-invasives sont disponibles pour l'évaluation de la fibrose hépatique et font l'objet de la Section 1.3.2 suivante.

1.3.2 Les tests non-invasifs

Afin de mettre en évidence une fibrose hépatique avancée chez des patients asymptomatiques, des méthodes non-invasives comprenant des tests sanguins ou d'élastométrie ont été développées. Contrairement à la biopsie, ces méthodes sont sans risque pour le patient et facilement répétables durant le suivi.

1.3.2.1 Tests sanguins

Les tests sanguins permettent de diagnostiquer la présence ou non de fibrose du foie dans les maladies hépatiques chroniques. Ils sont calculés à partir de biomarqueurs sanguins obtenus à l'aide d'une simple prise de sang : ils peuvent facilement être répétés dans le temps afin de contrôler une possible évolution de la maladie sans risque d'hémorragie ou autres complications pour le patient. De plus, ils sont largement accessibles puisqu'ils peuvent être théoriquement prescrits par tous les médecins.

Il existe deux types de tests sanguins : les tests simples et les tests spécialisés (BARANOVA et al. 2011). Les tests simples combinent des paramètres cliniques (comme l'âge ou le sexe du patient) et des marqueurs sanguins usuels peu coûteux (comme les transaminases ou les plaquettes). Leur calcul est accessible grâce à des sites internet ou applications gratuites contrairement aux tests sanguins spécialisés souvent payants. Les tests spécialisés comprennent des biomarqueurs directs de fibrose correspondants aux protéines constitutives de la fibrose hépatique et/ou aux enzymes impliquées dans les processus de fibrogénèse/fibrolyse intra-hépatique. Une partie de ces protéines passent dans le sang et leur taux sérique est bien corrélé aux stades de fibrose hépatique. Différents scores sanguins simples et spécialisés utilisés en pratique clinique sont présentés dans la Table 1.3.

Le score de Forns, l'APRI, le NFS, le FIB-4, et l'eLIFT sont des tests simples de fibrose, alors que le FibroTest, le FibroMètre, l'HEPASCORE, et l'ELF sont des tests spécialisés de fibrose. Le NFS a été conçu spécialement pour évaluer la fibrose dans la NAFLD. Le score de Forns ainsi que les tests APRI, FibroTest, FibroMètre et HEPASCORE ont été créés dans le cadre de l'hépatite C, tandis que le FIB-4 a été développé chez des patients ayant une co-infection VHC-VIH. Même si ces derniers scores ont été élaborés dans le cadre des hépatites virales, ils sont aussi performants dans la NAFLD. Enfin, les scores ELF et eLIFT ont été établis chez des patients atteints de maladies hépatiques chroniques de causes variées. Les tests sanguins les plus validés dans la NAFLD sont le FIB-4, le NFS et l'ELF, contre le FIB-4, le FibroMètre et le FibroTest pour l'hépatite C (EASL 2021).

TABLE 1.3 – Tests sanguins simples et spécialisés utilisés dans le cadre de la détection de fibrose hépatique

Tests simples	Marqueurs sanguins
Score de Forns (FORNS et al. 2002)	Âge, Cholesterol, GGT, Plaquettes
APRI (WAI et al. 2003)	ASAT, Plaquettes
FIB-4 (STERLING et al. 2006)	Âge, ALT, AST, Plaquettes
NFS (ANGULO et al. 2007)	Âge, Albumine, ALT, AST, Hyperglycémie IMC, Plaquettes
eLIFT (BOURSIER et al. 2017)	Âge, AST, GGT, Plaquettes, Sexe, TP
Tests spécialisés	Marqueurs sanguins
FibroTest (IMBERT-BISMUT et al. 2001)	α 2-macroglobuline, Apolipoprotéine A1, Âge, Bilirubine, GGT, Haptoglobine, Sexe
ELF (ROSENBERG et al. 2004)	Acide Hyaluronique, Âge, PIIIP, TIMP-1
FibroMètre (CALÈS et al. 2005)	α 2-macroglobuline, Acide Hyaluronique, Âge, AST, Plaquettes, Sexe, TP, Urée
HEPASCORE (ADAMS et al. 2005)	α 2-macroglobuline, Acide Hyaluronique, Âge, Bilirubine, GGT, Sexe

Abréviations : ALT = ALanine aminoTransférase ; APRI = *AST Platelet Ratio Index* ; AST = ASpartate aminoTransférase ; ELF = *Enhanced Liver Fibrosis* ; eLIFT = easy LIVER Fibrosis Test ; FIB-4 : FIBrosis-4 ; GGT = Gamma-GlutamylTransférase ; IMC = Indice de Masse Corporelle ; NFS = *NAFLD Fibrosis Score* ; TP = Taux de Prothrombine ; PIIIP = *amino-terminal Propeptide of type III Procollagen* ; TIMP-1 = *Tissue Inhibitor of matrix MetalloProteinase 1*.

Les avantages et inconvénients de chacun de ces tests ont été décrits par WONG 2018. Ils ont été développés pour le diagnostic de la fibrose avancée dans des cohortes de patients ayant subi une biopsie du foie dans des centres de soins tertiaires (BOURSIER et TSOCHATZIS 2021). Ils sont bien adaptés à ces populations : il faut donc être prudent lors de l'utilisation de ces tests lors de dépistage dans des populations moins spécifiques. De plus, des facteurs extra-hépatiques peuvent influencer la valeur de biomarqueurs inclus dans les tests sanguins et ainsi induire un risque de résultats faux-positifs ou faux-négatifs (POYNARD et al. 2004).

1.3.2.2 Élasticité hépatique

Une seconde méthode utilisée dans le diagnostic non-invasif de la fibrose hépatique est la mesure de l'élasticité (ou dureté) du foie. En pratique, il existe beaucoup d'appareils permettant de la mesurer : le « Fibroscan » (Echosens ; Paris, France) a été le premier commercialisé en 2004. Disponible dans les centres spécialisés, le Fibroscan est un examen

indolore non-invasif qui est très rapide à réaliser avec un résultat immédiat (CASTERA, FORNS et ALBERTI 2008). Il est utilisé dans toutes les maladies hépatiques chroniques et est schématisé en Figure 1.7.

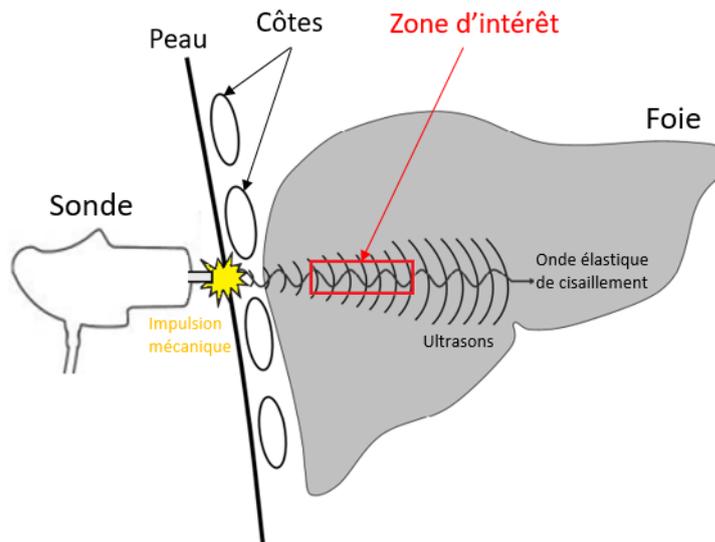


FIGURE 1.7 – Mesure de l'élasticité hépatique à l'aide du Fibroscan

Afin de réaliser un examen à l'aide du Fibroscan, le patient doit d'abord s'allonger en plaçant son bras droit derrière la tête afin de libérer l'hypochondre. Après avoir appliqué un gel, le clinicien pose une sonde sur la peau du patient en regard du foie. Elle envoie alors une impulsion mécanique, générant une onde élastique de cisaillement qui traverse l'organe. Un Doppler, inclus dans la sonde, mesure la vitesse de propagation de l'onde au sein d'une zone d'intérêt située entre 2.5 et 6.5 centimètres de profondeur sous la peau avec un diamètre d'un centimètre. Le volume hépatique étudié est 100 fois supérieur à celui d'une biopsie du foie (CASTERA, FORNS et ALBERTI 2008). L'élasticité hépatique (mesurée en kiloPascal - kPa) est enfin déduite de la vitesse de l'onde. Le principe est donc le suivant : plus la propagation de l'onde est rapide, plus le foie est dur, et donc fibreux.

Lors d'un Fibroscan, la machine réalise dix mesures et indique la médiane comme résultat de l'examen. Cependant, il est important de préciser qu'il peut exister une certaine variabilité des mesures liée au niveau de fibrose hépatique, à l'IMC, et au site de mesure (BOURSIER et al. 2008). Dans le cadre de la NAFLD par exemple, les patients sont

souvent en situation d'obésité avec une importante quantité de graisse abdominale qui influe sur la vitesse de propagation des ondes : le résultat est donc biaisé. Un Fibroscan ne sera donc pas considéré comme fiable si l'écart entre le minimum et le maximum des dix mesures réalisées est trop important. Une règle a été fixée, imposant que si l'écart interquartile (IQR : différence entre le troisième et le premier quartile des mesures) divisé par la médiane est strictement supérieur à 0.3 et que la médiane est supérieure ou égale à 8.0 kPa, alors l'examen effectué n'était pas fiable (BOURSIER et al. 2019). Un des inconvénients de cette méthode est que les seuils de fibrose sont mal définis au sein d'une même étiologie et diffèrent selon les maladies hépatiques.

De nombreux algorithmes permettant d'évaluer la fibrose hépatique et associant plusieurs tests non-invasifs ont été développés dans un premier temps dans l'hépatite C, puis plus récemment dans la NAFLD (CASTERA, FORNS et ALBERTI 2008; BOURSIER et al. 2016). Selon les recommandations de l'EASL, une combinaison séquentielle d'un test sanguin simple comme premier test, puis d'un Fibroscan a une meilleure précision diagnostique qu'un test seul (EASL 2021). Facilement accessibles et peu coûteux, les tests sanguins simples (voir Table 1.3) utilisés en première ligne permettent d'exclure les patients à faible risque de complications hépatiques liées à la fibrose et peuvent être utilisés en dehors des centres spécialisés en hépatologie. Le Fibroscan est utilisé dans un deuxième temps chez les malades classés à risque moyen/élevé par les tests sanguins simples. Il peut aussi être remplacé par des tests sanguins spécialisés : d'autres algorithmes combinant eLIFT/FibroMètre (BOURSIER et al. 2017) ou FIB-4/ELF (SRIVASTAVA et al. 2019b) ont également montré de bonnes performances. De plus, ces algorithmes non-invasifs ont l'avantage par rapport à la biopsie de pouvoir être répétés régulièrement durant le suivi du patient afin d'évaluer la progression de la fibrose hépatique.

1.4 Cohortes

Afin de répondre aux différents objectifs énoncés lors du préambule de cette thèse, trois cohortes sont utilisées : la première inclut des patients atteints de NAFLD, tandis que les deux suivantes contiennent des patients ayant contracté le VHC.

Cohorte NAFLD Cette cohorte regroupe 1376 patients atteints de NAFLD provenant de 5 centres tertiaires : Angers (France), Linköping et Stockholm (Suède), et les hôpitaux Virgen del Rocio et Valme de Séville (Espagne). Les patients ont été inclus de manière

rétrospective, à l'exception de la cohorte Linköping où les malades ont été inclus prospectivement. Pour chacun d'entre eux, la NAFLD a été diagnostiquée par la présence d'une stéatose hépatique à l'imagerie ou à la biopsie en excluant les causes suivantes : traitement induisant une stéatose, consommation excessive d'alcool ou autres maladies hépatiques chroniques. Les patients inclus présentaient au moins un test sanguin FIB-4 et/ou un Fibroscan. Lorsque des données de biopsie hépatique étaient disponibles, les lésions histologiques ont été évaluées selon la classification du NASH CRN vue en Section 1.3.1. Durant le suivi, les événements cliniques observés étaient les complications de cirrhose (ascite, infection du liquide d'ascite, syndrome hépato-rénal, varice gastro-œsophagienne, hémorragie digestive par rupture de varice, insuffisance hépatique et encéphalopathie), le CHC et le décès. Cette cohorte a été utilisée lors des Chapitres 3 et 5.

Cohorte ANRS CO12 CirVir C'est une cohorte prospective financée par l'Agence Nationale de la Recherche (ANRS). Elle inclut 1323 patients présentant une cirrhose prouvée par biopsie due au VHC entre mars 2006 et juin 2012 dans 35 centres de soins français spécialisés en hépatologie (TRINCHET et al. 2015). Les patients ont été suivis par les médecins tous les six mois environ selon les recommandations de surveillance de la cirrhose. Chaque visite donnait lieu à un recueil des données cliniques, biologiques et échographiques. Les différents événements hépatiques étudiés durant le suivi étaient les complications de cirrhose (ascite, encéphalopathie, hémorragie par rupture de varice gastro-œsophagienne et ictère), le CHC et le décès. Le diagnostic du CHC a été établi soit par un examen histologique, soit par des examens radiologiques selon les recommandations en cours. La cohorte ANRS CO12 CirVir a été exploitée lors du Chapitre 4.

Cohorte ANRS CO22 HEPATHER Cette seconde cohorte prospective de l'ANRS est constituée de 10412 patients issus de 32 centres hépatologiques français depuis août 2012 et ayant été infectés par le VHC (CARRAT et al. 2019). Le suivi des malades se fait une fois par an avec collecte de données de la pratique clinique (données biologiques principalement). Les mêmes événements hépatiques que pour la cohorte ANRS CO12 CirVir sont enregistrés sur des formulaires de données spécifiques (complications de cirrhose, CHC, décès). Cette cohorte étant basée sur une étude d'observation, le dépistage du CHC ou de la progression de la fibrose étaient pris en charge par le médecin selon les recommandations de l'EASL (EASL 2018). C'est également lors du Chapitre 4 que la cohorte ANRS CO22 HEPATHER a été utilisée.

ÉTAT DE L'ART STATISTIQUE

Ce chapitre propose un état des lieux de différents modèles utilisés dans le domaine des biostatistiques permettant l'analyse de données longitudinales, l'analyse de données de survie, ainsi que l'analyse conjointe de ces deux derniers types de données. Des critères de performance prédictive de ces différents modèles y sont également présentés, ainsi qu'une introduction à la multiplicité des tests.

2.1 Modèle linéaire généralisé à effets mixtes

Un modèle de régression classique suppose l'indépendance entre toutes les observations. Dans le cas de données longitudinales, cette hypothèse reste valable pour les données inter-individu lorsque les sujets sont sélectionnés au hasard et individuellement dans la population. Cependant, deux observations du même sujet sont corrélées par le partage des mêmes caractéristiques rendant ainsi les modèles de régression classiques inexploitable.

Les modèles à effets mixtes, introduits par MCCULLAGH et NELDER 1989, sont ainsi nécessaires afin de tenir compte de la corrélation intra-individu. Pour cela, les modèles à effets mixtes incluent des effets fixes communs pour tous les individus, des effets aléatoires propres à chaque individu, et des erreurs de mesure. Les effets fixes expliquent la variation ou la distribution de la variable à expliquer, tandis que les effets aléatoires traduisent la variabilité liée à chaque sujet (la variabilité intra-individu).

Plusieurs types de modèles à effets mixtes existent. Nous introduirons d'une part les modèles linéaires à effets mixtes qui sont un cas particulier des modèles linéaires généralisés à effets mixtes, puis les modèles mixtes à classes latentes.

2.1.1 Modèle linéaire à effets mixtes

Soit $\mathbb{Y}_N = (Y_1, \dots, Y_N)$ un N -échantillon gaussien tel que, pour tout $i \in \{1, \dots, N\}$,

$$Y_i = \left(Y_i^{(1)}, \dots, Y_i^{(n_i)} \right)^\top \in \mathbb{R}^{n_i},$$

où $n_i \in \mathbb{N}^*$ est le nombre de mesures observées pour l'individu i , et $^\top$ désigne la transposée. Ainsi, pour tout $j \in \{1, \dots, n_i\}$, $Y_i^{(j)}$ représente la j -ème mesure associée à l'individu i . Chaque observation $Y_i^{(j)}$ n'est qu'une mesure bruitée de sa vraie valeur qui reste inobservée. L'idée du modèle linéaire à effets mixtes (HENDERSON et al. 1959 ; VERBEKE et MOLENBERGHS 2009) est d'expliquer $Y_i^{(j)}$ à la fois par des facteurs à effets fixes et par des facteurs à effets aléatoires prenant ainsi en compte la corrélation intra-individuelle et l'hétérogénéité entre les individus. Les effets fixes interviennent au niveau de la moyenne du modèle, tandis que les effets aléatoires agissent au niveau de la variance du modèle.

Pour tout $i \in \{1, \dots, N\}$, pour tout $j \in \{1, \dots, n_i\}$, le modèle linéaire à effets mixtes s'écrit

$$Y_i^{(j)} = X_{i,j}^\top \beta + Z_{i,j}^\top b_i + \epsilon_i^{(j)},$$

où

- $Y_i^{(j)}$ est la variable aléatoire réponse associée à la j -ème mesure de l'individu i ,
- $\beta \in \mathbb{R}^{p+1}$, $p \in \mathbb{N}^*$, est le vecteur des effets fixes associé au vecteur de covariables $X_{i,j}$, représentant la part expliquée de $Y_i^{(j)}$ par les variables explicatives communes à tous les individus,
- $b_i \in \mathbb{R}^q$, $q \in \mathbb{N}^*$, $q \leq p$, est le vecteur des effets aléatoires associé au vecteur de covariables $Z_{i,j}$, où $Z_{i,j}$ est un sous-vecteur de $X_{i,j}$, traduisant à la fois l'écart à la moyenne attendue de $Y_i^{(j)}$ pour le sujet i et la corrélation entre les mesures répétées de l'individu i . Les b_i sont des vecteurs gaussiens indépendants et identiquement distribués (i.i.d.) centrés de matrice de variance-covariance B ,
- $\epsilon_i^{(j)}$ est un résidu gaussien centré de variance $\sigma^2 > 0$. De plus, $\epsilon_i^{(1)}, \dots, \epsilon_i^{(n_i)}$ sont indépendants, et $\epsilon_i^{(j)}$ est indépendant de b_i .

Le vecteur $X_{i,j}$ est de taille $p+1$ mais ne prend en compte que p covariables puisqu'un intercept a été ajouté au modèle. Comme il est expliqué dans l'Exemple 1, il est possible d'intégrer des intercepts ou pentes aléatoires dans le modèle.

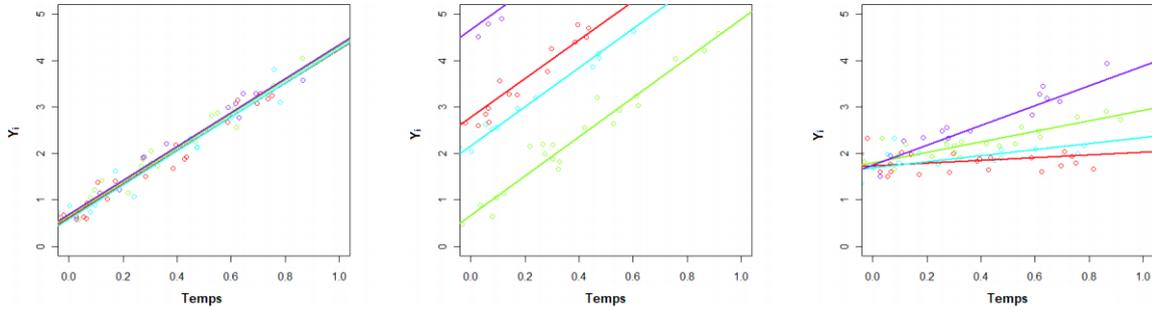


FIGURE 2.1 – À gauche : modèle sans effet aléatoire ; au centre : modèle avec un intercept aléatoire ; à droite : modèle avec une pente aléatoire

Exemple 1. Sur la Figure 2.1, chaque couleur correspond à une simulation. Sans intercept ni pente aléatoires, les points et droites de régression se superposent sur la figure de gauche. L'intégration d'un intercept aléatoire sur la figure au centre permet d'avoir l'ordonnée à l'origine différente selon chaque simulation. La figure de droite illustre l'effet de la pente aléatoire : à chaque simulation, le coefficient directeur de la droite de régression diffère.

La matrice de variance-covariance des effets aléatoires B peut prendre différentes structures : deux différentes sont présentées dans l'Exemple 2.

Exemple 2. La Figure 2.2 présente deux types de matrices de variance-covariance de taille $q = 5$. Soit $\varpi_{..} > 0$.

$$\begin{pmatrix} \varpi_{11} & \varpi_{12} & \varpi_{13} & \varpi_{14} & \varpi_{15} \\ \varpi_{12} & \varpi_{22} & \varpi_{23} & \varpi_{24} & \varpi_{25} \\ \varpi_{13} & \varpi_{23} & \varpi_{33} & \varpi_{34} & \varpi_{35} \\ \varpi_{14} & \varpi_{24} & \varpi_{34} & \varpi_{44} & \varpi_{45} \\ \varpi_{15} & \varpi_{25} & \varpi_{35} & \varpi_{45} & \varpi_{55} \end{pmatrix} \quad \begin{pmatrix} \varpi_{11}^2 & 0 & 0 & 0 & 0 \\ 0 & \varpi_{22}^2 & 0 & 0 & 0 \\ 0 & 0 & \varpi_{33}^2 & 0 & 0 \\ 0 & 0 & 0 & \varpi_{44}^2 & 0 \\ 0 & 0 & 0 & 0 & \varpi_{55}^2 \end{pmatrix}$$

FIGURE 2.2 – À gauche : matrice de variance-covariance non-structurée ; à droite : matrice de variance-covariance diagonale

La structure de la matrice de gauche est non-structurée. Elle est utilisée lorsqu'aucune hypothèse n'est faite sur la structure de dépendance des effets aléatoires : elles sont potentiellement toutes corrélées entre elles. Tous les coefficients de la matrice peuvent être

non-nuls et il faudra alors estimer $\frac{q(q+1)}{2}$ paramètres. La matrice de droite est quant à elle diagonale. Les covariables ne sont pas corrélées entre elles et la matrice contient seulement les variances de chacune où q coefficients sont à estimer.

En pratique, nous commencerons l’analyse en faisant le moins d’hypothèses possibles sur les effets aléatoires et la matrice de variance-covariance B : le type non-structuré sera privilégié. Cependant, cela donnera parfois des modèles trop complexes qui ne peuvent converger. Dans ce cas, nous supposerons que B est une matrice diagonale : les effets aléatoires sont indépendants les uns des autres.

Pour tout $i \in \{1, \dots, N\}$, l’espérance et la variance marginales de $Y = (Y_i)_{1 \leq i \leq N}$ sont données par

$$\mathbb{E}[Y_i] = X^i \beta \text{ et } \mathbb{V}(Y_i) = Z^i B Z^{iT} + \sigma^2 I_{n_i},$$

où X^i et Z^i sont les matrices dont les lignes sont respectivement $X_{i_j}^T$ et $Z_{i_j}^T$, et I_{n_i} la matrice identité de taille $n_i \times n_i$. L’espérance marginale représente la valeur moyenne de Y_i pour tous les individus qui partagent les mêmes caractéristiques X^i que le i -ème individu.

De plus, pour tout $i \in \{1, \dots, N\}$, l’espérance de Y_i conditionnellement aux effets aléatoires vaut

$$\mathbb{E}[Y_i | b_i] = X^i \beta + Z^i b_i.$$

Elle est représentée par l’espérance spécifique à chaque individu et la tendance aléatoire expliquée par les effets aléatoires.

Le vecteur des effets fixes β est le même pour tous les individus, contrairement au vecteur des effets aléatoires b_i qui est spécifique à chacun. Ces vecteurs sont inconnus et sont à estimer afin de pouvoir exploiter nos données et ainsi pouvoir en tirer des informations. Les modèles à effets mixtes étant un cas particulier des modèles linéaires généralisés (MCCULLAGH et NELDER 1989), la vraisemblance et l’estimation des coefficients sont présentés dans la Section 2.1.2 suivante.

Lorsque la variable réponse Y ne suit pas une loi normale, certaines transformations comme celles de type Box-Cox (BOX et COX 1964) permettent de se ramener à une loi normale. Cela apporte un gain de puissance et d’interprétation au modèle lorsque c’est possible. Si ce n’est pas le cas, le modèle linéaire à effets mixtes ne peut plus être utilisé

et il faut recourir au modèle linéaire généralisé à effets mixtes (MCCULLAGH et NELDER 1989).

2.1.2 Modèle linéaire généralisé à effets mixtes (GLMM)

Les modèles linéaires généralisés à effets mixtes peuvent être utilisés lorsque la variable à expliquer est une variable binaire, une variable de dénombrement, ou comme mentionné précédemment, une variable non-gaussienne. L'idée de cette généralisation est d'expliquer l'espérance de la variable réponse conditionnellement aux covariables par un prédicteur linéaire via une « fonction de lien ».

2.1.2.1 Définition du modèle linéaire généralisé à effets mixtes

Pour tout $i \in \{1, \dots, N\}$, pour tout $j \in \{1, \dots, n_i\}$, le modèle linéaire généralisé à effets mixtes est défini par :

- $Y_i^{(j)}$ la variable aléatoire réponse associée à la mesure j de l'individu i dont la densité conditionnellement aux effets aléatoires $f_{Y_i^{(j)}|b_i}$ appartient à une famille exponentielle et s'écrit

$$f_{Y_i^{(j)}|b_i}(y_i^{(j)}|b_i; \psi_i^{(j)}, \phi) = \exp \left[\frac{y_i^{(j)} \psi_i^{(j)}(b_i) - C(\psi_i^{(j)}(b_i))}{a(\phi)} - d(y_i^{(j)}, \phi) \right],$$

où $C(\cdot)$, $d(\cdot, \cdot)$ et $a(\cdot)$ sont des fonctions connues, et $\psi_i^{(j)}(b_i)$ et ϕ sont respectivement les paramètres canoniques et de dispersion,

- $\eta_i^{(j)} = X_{i_j}^\top \beta + Z_{i_j}^\top b_i$ le prédicteur linéaire associé à la mesure j de l'individu i , avec $\beta \in \mathbb{R}^{(p+1)}$ le vecteur des effets fixes associé au vecteur de covariables $X_{i_j} \in \mathbb{R}^{(p+1)}$ et $b_i \in \mathbb{R}^q$ le vecteur des effets aléatoires associé au vecteur de covariables $Z_{i_j}^\top \in \mathbb{R}^q$,
- une « fonction de lien » notée $g : \mathbb{R} \rightarrow \mathbb{R}$, telle que

$$g(\mathbb{E}[Y_i^{(j)}|b_i]) = X_{i_j}^\top \beta + Z_{i_j}^\top b_i.$$

2.1.2.2 Estimation des paramètres

L'ensemble des différents paramètres du modèle θ_{GLMM} (comprenant $\psi_i^{(j)}(b_i)$ et ϕ) est estimé par la méthode du maximum de vraisemblance. Conditionnellement aux effets

aléatoires, les différentes mesures d’un même individu sont indépendantes entre elles. Ainsi, pour tout $i \in \{1, \dots, N\}$,

$$f_{Y|b}(y_i|b_i; \theta_{\text{GLMM}}) = \prod_{j=1}^{n_i} f_{Y|b}(y_i^{(j)}|b_i; \theta_{\text{GLMM}}), \quad (2.1)$$

et en utilisant l’indépendance entre les individus, la vraisemblance vaut

$$L_{\text{GLMM}}(\theta_{\text{GLMM}}|\mathbb{Y}_N) = \prod_{i=1}^N f_Y(y_i; \theta_{\text{GLMM}}). \quad (2.2)$$

En conditionnant par les effets aléatoires, (2.2) se réécrit

$$\prod_{i=1}^N \int_{\mathbb{R}^q} f_{Y|b}(y_i|b_i; \theta_{\text{GLMM}}) f_b(b_i; \theta_{\text{GLMM}}) db_i,$$

où $f_b(b_i; \theta_{\text{GLMM}})$ est la densité associée aux effets aléatoires. Enfin, en utilisant (2.1), la vraisemblance vaut

$$\prod_{i=1}^N \int_{\mathbb{R}^q} \prod_{j=1}^{n_i} f_{Y|b}(y_i^{(j)}|b_i; \theta_{\text{GLMM}}) f_b(b_i; \theta_{\text{GLMM}}) db_i.$$

Le maximum de la log-vraisemblance est obtenu à l’aide d’algorithmes d’optimisation tels que l’algorithme de Newton-Rapson (YPMA 1995), ou encore l’algorithme d’*Expectation-Maximisation* (DEMPSTER, LAIRD et RUBIN 1977). Enfin, les estimations des covariances relatives aux estimations conjointes des différents paramètres sont obtenues en calculant l’inverse de la matrice d’information de Fisher en $\hat{\theta}_{\text{GLMM}}$.

Remarque 1. *Dans ce type de modèle, nous faisons l’hypothèse de départ que les individus sont indépendants entre eux. Cependant, quand des individus proviennent d’un même groupe, cette hypothèse d’indépendance peut être mise à mal et ne pas être respectée. En général, pour traiter ce cas, il suffit d’inclure la variable de groupe aux effets aléatoires car cette caractéristique d’appartenance à tel ou tel groupe sera propre au sujet. Par exemple, dans un jeu de données contenant des patients issus de différents centres hospitaliers, la variable « Centre hospitalier » sera considérée en tant qu’effet aléatoire.*

En résumé, les modèles à effets mixtes permettent de tenir compte d'une corrélation intra-individu. Contrairement aux modèles de régression ordinaires, ils intègrent des effets aléatoires spécifiques à chaque individu en plus des effets fixes. Ils peuvent donc tenir compte de l'hétérogénéité de la variance parmi les réponses de différents individus, ainsi que du modèle de covariance parmi les réponses au sein d'un même individu.

Les modèles à effets mixtes supposent la population homogène : les individus possèdent les mêmes caractéristiques et la variable réponse varie autour d'un unique profil moyen. Cependant, cette hypothèse peut ne pas être réaliste dans certaines situations. Un moyen de s'en affranchir est de considérer des modèles mixtes à classes latentes.

2.1.3 Modèle mixte à classes latentes

Les modèles mixtes à classes latentes permettent de considérer différents profils moyens au sein de la population. Chaque groupe d'individus partageant les mêmes caractéristiques est représenté par une classe latente. Ainsi, nous pouvons déterminer les différentes trajectoires moyennes de la variable que l'on souhaite expliquer selon chaque groupe. Ces modèles se présentent sous la même forme que les modèles à effets mixtes précédents (cf Section 2.1.2) combinés à un modèle logistique multinomial (HAUSMAN et MCFADDEN 1984). Ils ont été mis en place par VERBEKE et LESAFFRE 1996, puis développés par MUTHÉN et SHEDDEN 1999.

2.1.3.1 Définition du modèle mixte à classes latentes

On suppose qu'il existe G groupes d'individus au sein de la population, définissant ainsi les G classes latentes. La variable discrète latente c_i permet d'indiquer à quelle classe $g \in \{1, \dots, G\}$ appartient l'individu $i \in \{1, \dots, N\}$. Chaque individu appartient à une unique classe. Le modèle logistique multinomial inclus dans le modèle mixte à classes latentes permet de déterminer la probabilité pour un individu i d'appartenir à une certaine classe g .

Pour tout $(i, j, g) \in \{1, \dots, N\} \times \{1, \dots, n_i\} \times \{1, \dots, G\}$, le modèle mixte à classes latentes s'écrit

$$\begin{cases} \pi_{ig} &= \mathbb{P}(c_i = g | X_i^{\text{mlt}}) = \frac{\exp(\xi_{0g} + X_i^{\text{mlt}\top} \xi_{1g})}{\sum_{l=1}^G \exp(\xi_{0l} + X_i^{\text{mlt}\top} \xi_{1l})}, & \xi_{0G} = 0, \quad \xi_{1G} = 0, \\ Y_i^{(j)} &= X_{i_j}^\top \beta + X_{i_j}^{\text{cl}\top} \gamma_g + Z_{i_j}^\top b_i + \epsilon_i^{(j)}, \end{cases}$$

avec

- π_{ig} la probabilité que l'individu i appartienne à la classe g selon le modèle logistique multinomial expliqué par le vecteur de covariables X_i^{mlt} . La classe G est fixée comme étant la classe de référence, d'où $\xi_{0G} = 0$ et $\xi_{1G} = 0$, afin de s'assurer de l'identifiabilité du modèle,
- β le vecteur des effets fixes commun à toutes les classes et associé au vecteur de covariables X_{i_j} ,
- γ_g le vecteur des effets fixes associé au vecteur de covariables $X_{i_j}^{\text{cl}}$ représentant la part expliquée de $Y_i^{(j)}$ par les variables explicatives spécifiques à chaque classe,
- b_i le vecteur des effets aléatoires associé au vecteur de covariables Z_{i_j} tel que $b_i \sim \mathcal{N}(\mu_g, B_g)$,
- $\epsilon_i = (\epsilon_i^{(j)})_{1 \leq j \leq n_i} \sim \mathcal{N}(0, \Sigma_i)$ tel que $\epsilon_i^{(j)}$ est indépendant de $\epsilon_i^{(j')}$ pour tout $j \neq j'$, $j, j' \in \{1, \dots, n_i\}$.

2.1.3.2 Estimation des paramètres

L'estimation des paramètres pour les modèles mixtes à classes latentes se fait par maximum de vraisemblance. Soient $g \in \{1, \dots, G\}$ et θ_G l'ensemble des paramètres à estimer. Pour tout $i \in \{1, \dots, N\}$, la distribution marginale de Y_i conditionnellement à la classe latente g est

$$\mathcal{N}(X^i \beta + X^{\text{cl}i} \gamma_g + Z^i \mu_g, Z^i B_g Z^{i\top} + \Sigma_i),$$

où X^i , $X^{\text{cl}i}$ et Z^i sont les matrices dont les lignes sont respectivement $X_{i_j}^\top$, $X_{i_j}^{\text{cl}\top}$ et $Z_{i_j}^\top$. En utilisant l'indépendance inter-individu, la vraisemblance pour G fixé s'écrit :

$$\prod_{i=1}^N \left[\sum_{g=1}^G \mathbb{P}(c_i = g | X^{\text{mlt}i}; \theta_G) \times f_{Y|c}(Y_i | c_i = g; X^i, X^{\text{cl}i}, Z^i; \theta_G) \right],$$

où $f_{Y|c}$ désigne la fonction de densité de la loi normale multivariée d'espérance et de variance respectivement

$$\mathbb{E} [f_{Y|c}] = X^i \beta + X^{cl^i} \gamma_g + Z^i \mu_g \text{ et } \mathbb{V} [f_{Y|c}] = Z^i B_g Z^{i\top} + \Sigma_i.$$

Afin de déterminer le nombre optimal de classes latentes, il convient de retenir le modèle minimisant le *Bayesian Information Criterion* (BIC) (SCHWARZ 1978) défini par

$$-2 \times \ln(L) + k \times \ln(N),$$

où L est la vraisemblance du modèle estimé, N le nombre d'individus, et k le nombre de paramètres du modèle à estimer. En pratique, il est préféré à l'*Akaike information criterion* (AKAIKE 1974) puisqu'il privilégie des modèles plus parcimonieux en pénalisant davantage les modèles.

2.2 Analyse de survie : généralités

Depuis une cinquantaine d'années, les statisticiens ont développé de nombreux modèles rendant possible l'analyse de données issues d'essais cliniques (VAN HOUWELINGEN et PUTTER 2011). Par exemple, ces modèles permettent de fournir au patient une estimation de la durée de survie après le diagnostic d'une maladie : l'espérance de vie d'une maladie grave peut être de quelques mois, ou de quelques années pour une maladie moins grave.

Les données de survie sont utilisées pour analyser le délai écoulé avant l'apparition d'un évènement d'intérêt. L'expression « temps de survie » désigne la période pendant laquelle l'individu est observé jusqu'à cet évènement (ou potentiellement la fin de l'étude s'il n'arrive pas avant). En pratique, l'évènement d'intérêt n'est pas forcément la mort. Le but est alors d'étudier la probabilité de ne pas déclarer un évènement jusqu'à un instant t , et de voir comment certaines variables explicatives peuvent modifier cette probabilité.

2.2.1 Fonctions associées aux distributions de survie

Soit T une variable aléatoire continue positive ou nulle, représentant la durée de survie avant l'évènement d'intérêt. La loi de probabilité de T est caractérisée par cinq fonctions définies sur \mathbb{R}^+ qui sont toutes liées entre elles.

Soit $t \in \mathbb{R}^+$ fixé. La fonction de survie au temps t , notée $S(t)$, représente la probabilité que le temps de survie d'un individu soit supérieur à t . Autrement dit, c'est la probabilité que l'évènement ne se produise pas avant t . C'est une fonction décroissante définie pour tout $t \geq 0$ par

$$S(t) = \begin{cases} 1 & \text{si } t = 0 \\ \mathbb{P}(T > t) & \text{si } t > 0 \end{cases}.$$

La fonction de répartition au temps t , notée $F(t)$, correspond à la probabilité de connaître l'évènement durant la période $[0, t]$. Elle est définie pour tout $t \geq 0$ par

$$F(t) = \mathbb{P}(T \leq t) = 1 - S(t).$$

Soit $\Delta t \in \mathbb{R}^+$ fixé. La fonction de densité $f(t)$ est égale à la probabilité de réalisation de l'évènement entre les temps t et $(t + \Delta t)$. Pour tout $t \geq 0$, elle s'écrit

$$\begin{aligned} f(t) &= \lim_{\Delta t \rightarrow 0^+} \frac{\mathbb{P}(t \leq T \leq t + \Delta t)}{\Delta t} \\ &= \lim_{\Delta t \rightarrow 0^+} \frac{\mathbb{P}(T \leq t + \Delta t) - \mathbb{P}(T \leq t)}{\Delta t} \\ &= \lim_{\Delta t \rightarrow 0^+} \frac{F(t + \Delta t) - F(t)}{\Delta t} \\ &= F'(t) \\ &= -S'(t). \end{aligned}$$

La fonction de risque à l'instant t , notée $\lambda(t)$, représente la probabilité de subir l'évènement entre les instants t et $(t + \Delta t)$, tout en sachant que l'individu est encore considéré à risque au temps t . Pour tout $t \geq 0$, nous avons

$$\begin{aligned} \lambda(t) &= \lim_{\Delta t \rightarrow 0^+} \frac{\mathbb{P}(t \leq T \leq t + \Delta t \mid T \geq t)}{\Delta t} \\ &= \frac{1}{\mathbb{P}(t \leq T)} \lim_{\Delta t \rightarrow 0^+} \frac{\mathbb{P}(t \leq T \leq t + \Delta t)}{\Delta t} \\ &= \frac{f(t)}{S(t)} \\ &= -[\ln(S(t))]' . \end{aligned}$$

Enfin, la fonction de risque cumulé $\Lambda(t)$ correspond au cumul des risques instantanés

$\lambda(t)$ sur la période $[0, t]$. Pour tout $t \geq 0$,

$$\Lambda(t) = \int_0^t \lambda(u) \, du = \int_0^t -[\ln(S(u))]' \, du = -\ln(S(t)).$$

En pratique, lors d'une étude de survie, la variable aléatoire T peut n'être observée que partiellement. Nous détaillons par la suite, trois causes d'observation partielle usuellement rencontrées dans la littérature.

2.2.2 Troncature et censures

Afin d'illustrer les différentes causes d'observation partielle des données, nous nous appuyerons sur la Figure 2.3 de l'Exemple 3.

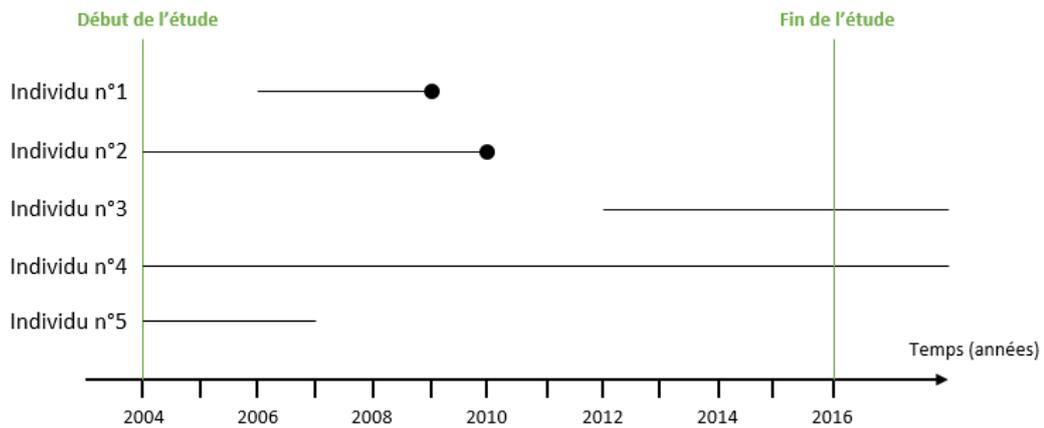


FIGURE 2.3 – Suivi de cinq patients entre 2004 et 2016

Exemple 3. La Figure 2.3 présente le suivi de cinq individus entre 2004 et 2016. Les points noirs désignent la survenue d'un évènement. Deux individus ont subi un évènement. L'individu 1 est entré dans l'étude en 2006, et a connu un évènement en 2009 ; tandis que l'individu 2 intégré dès le début de l'étude en 2004, a connu un évènement en 2010.

2.2.2.1 Troncature à gauche

La variable aléatoire T est tronquée à gauche par la variable aléatoire T_g lorsque T n'est observable que si

$$T > T_g.$$

Exemple 4. *Les données de la Figure 2.3 sont tronquées à gauche : l’inclusion à l’étude correspond à la première visite. Les individus sont considérés à risque de développer un évènement dès l’instant où ils commencent à être suivis, et non dès leur naissance. En pratique, la troncature à gauche est quasiment systématique.*

2.2.2.2 Censure à droite

Les données sont dites censurées à droite s’il existe au moins un individu n’ayant pas connu l’évènement au moment de sa dernière observation. Il y a principalement deux raisons à cette censure. Un individu censuré à droite peut ne pas avoir subi l’évènement avant la fin de l’étude : on parle d’individu « exclu vivant ». L’individu peut aussi quitter l’étude sans avoir connu l’évènement, soit parce qu’il ne veut plus être suivi, soit parce qu’il ne respecte plus les hypothèses d’inclusion à l’étude. Dans ce cas, on parle d’individu « perdu de vue ».

Soit C_d la variable aléatoire représentant le temps de censure à droite défini comme le temps entre l’inclusion à l’étude et le moment de la censure. Soit (\tilde{T}, δ) le couple de variables aléatoires où

$$\tilde{T} = \min(T, C_d)$$

représente la durée réellement observée, et

$$\delta = \mathbb{1}_{\{T \leq C_d\}}$$

l’indicateur d’évènement. Si l’évènement a eu lieu, alors $\delta = 1$; tandis que si l’évènement n’a pas été réalisé durant la durée d’observation, $\delta = 0$.

Exemple 5. *Sur la Figure 2.3, certaines données sont censurées à droite. L’individu 3 est entré dans l’étude en 2012, mais n’a pas connu d’évènement avant la fin en 2016 : il fait donc partie des individus exclus vivants. Les individus 4 et 5 ont tous les deux été intégrés dès le début de l’étude en 2004 et sont censurés à droite : le 4 parce qu’il n’a pas connu d’évènement avant la fin de l’étude (exclu vivant), le 5 parce qu’il en a été retiré en 2007 (perdu de vue).*

2.2.2.3 Censure par intervalle

La censure par intervalle intervient lorsqu'il est impossible de mesurer précisément la date de survenue de l'évènement : seul un intervalle de temps durant lequel s'est déroulé l'évènement est disponible. En pratique, il est fréquent que ce type de censure survienne. En effet, si les patients sont suivis de manière discontinue et que l'évènement d'intérêt est l'apparition d'une maladie, le diagnostic est défini le jour de la dernière visite mais le patient est tombé malade entre les deux dernières.

Dans tous les cas, il faut veiller à ce que l'indépendance entre censure ou troncature et l'évènement étudié soit respectée. Dans le cas contraire, le taux de survie serait biaisé.

2.2.3 Courbe de survie

Cette partie présente la courbe représentative de la fonction de survie

$$S : \mathbb{R}^+ \longrightarrow [0, 1], \quad t \longmapsto \mathbb{P}(T \geq t),$$

appelée courbe de survie. Une façon d'estimer cette courbe y est proposée, ainsi qu'un test statistique permettant d'en comparer plusieurs entre elles.

2.2.3.1 Estimateur de Kaplan-Meier

L'estimateur de Kaplan-Meier est un estimateur non-paramétrique utilisé pour estimer une courbe de survie lorsqu'aucune hypothèse n'est faite a priori sur la forme paramétrique de la loi de T (KAPLAN et MEIER 1958). C'est l'un des estimateurs les plus exploités en pratique.

On observe T sur la période $[0; t_f]$, où $t_f > 0$ et $t_1 \leq \dots \leq t_k \leq t_f$ est la discrétisation de l'intervalle de temps d'observation $[0; t_f]$ correspondant aux durées observées jusqu'à chaque survenue d'un évènement dans l'échantillon observé.

Pour tout $j \in \{1, \dots, k\}$, par définition

$$S(t_j) = \mathbb{P}(T > t_j).$$

Comme $\{T > t_j\} \supseteq \{T > t_{j-1}\}$, il vient que

$$\begin{aligned} S(t_j) &= \mathbb{P}(\{T > t_j\} \cap \{T > t_{j-1}\}) \\ &= \mathbb{P}(T > t_j | T > t_{j-1}) \mathbb{P}(T > t_{j-1}) \\ &= \prod_{i=1}^j \mathbb{P}(T > t_i | T > t_{i-1}) \\ &= \prod_{i=1}^j (1 - q_i), \end{aligned}$$

où $q_i = \frac{\mathbb{P}(t_{i-1} \leq T \leq t_i)}{\mathbb{P}(T > t_{i-1})}$ représente la probabilité de connaître l'évènement durant l'intervalle de temps $]t_{i-1}; t_i]$ sachant qu'on ne l'a pas connu avant l'instant t_{i-1} .

L'estimateur de Kaplan-Meier consiste à approximer q_i par \hat{q}_i défini par

$$\hat{q}_i = \frac{M_i}{R_i},$$

avec M_i le nombre d'individus ayant connu l'évènement à t_i , et R_i le nombre d'individus à risque au temps t_i n'ayant ni connu l'évènement, ni été censurés avant t_i .

Ainsi, la fonction de survie estimée \hat{S}_f associée à l'estimateur de Kaplan-Meier est

$$\hat{S}_f(t) = \prod_{i:t_i \leq t} \left(1 - \frac{M_i}{R_i}\right).$$

C'est une fonction en escalier décroissante continue à droite. Chaque saut représente l'apparition d'un ou plusieurs évènements au moment t pour l'échantillon de population concerné.

Exemple 6. La Figure 2.4 présente trois courbes de survie qui décroissent de différentes façons : les probabilités de survie selon les groupes sont différentes. La courbe du groupe 1 est quasiment plane, ce qui indique que ce groupe présente de faibles chances de survenue de l'évènement. À contrario, la courbe du groupe 3 décroît rapidement : les individus de ce groupe sont considérés comme plus à risque de subir un évènement. Le groupe 2 semble être un groupe intermédiaire en étant moins à risque que le groupe 3, mais plus que le groupe 1.

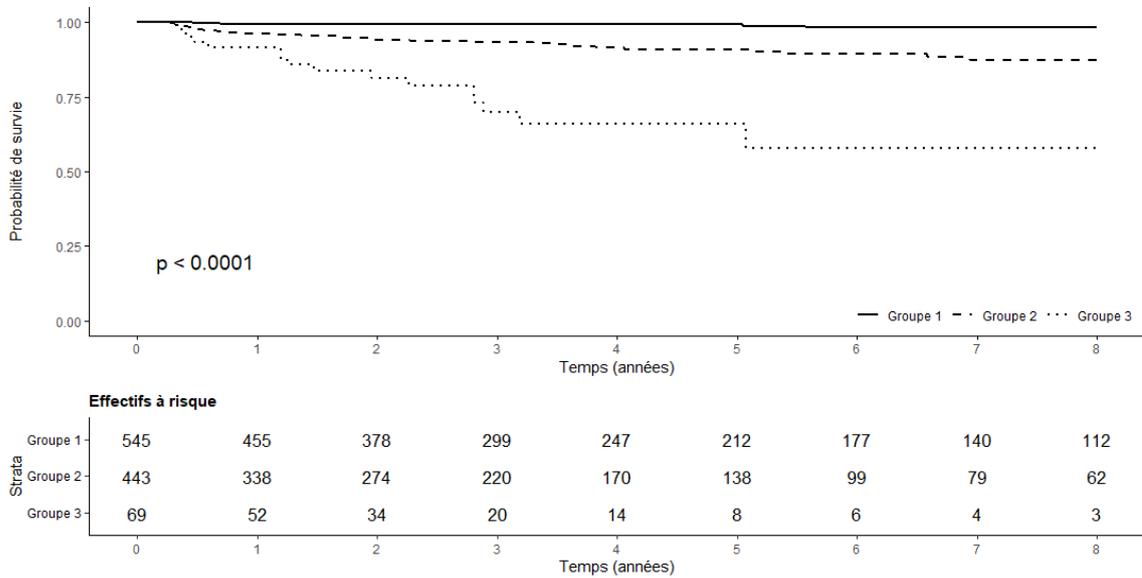


FIGURE 2.4 – Courbes de survie obtenues à l’aide de l’estimateur de Kaplan-Meier

La table en bas de la Figure 2.4 indique les effectifs à risque entre le début de l’observation $t = 0$ et $t = 8$ ans après. Au temps $t = 0$, on trouve les effectifs initiaux de chaque groupe : 545 individus dans le groupe 1, 443 dans le 2, et 52 dans le 3. À $t = 4$ ans, ces effectifs ont diminué. Des individus ont été censurés ou ont subi un évènement. On trouve alors 247 individus encore à risque dans le groupe 1, 170 dans le groupe 2 et 14 dans le dernier groupe à $t = 4$ ans.

La p-valeur $p < 0.0001$ présente sur la Figure 2.4 correspond à celle obtenue à l’aide du test du Logrank présenté dans la Section 2.2.3.2 suivante.

2.2.3.2 Test du Logrank

Les courbes de survie permettent une comparaison visuelle de la survie entre différents groupes d’individus, mais ne permettent pas de conclure à une différence significative du taux de survie entre ces groupes. Pour cela, le test du Logrank (MANTEL 1966 ; PETO et PETO 1972) est utilisé en pratique pour comparer $m \geq 2$ courbes de survie entre elles. Son fonctionnement repose sur la comparaison entre le nombre d’évènements observés et le nombre d’évènements attendus sous l’hypothèse nulle \mathcal{H}_0 d’égalité des fonctions de survie. Il teste

$$\mathcal{H}_0 : \ll S_1(t) = S_2(t) = \dots = S_m(t) \gg$$

contre

$$\mathcal{H}_1 : \ll \exists i, j \in \{1, \dots, m\} \text{ et } i \neq j, S_i(t) \neq S_j(t) \gg.$$

Une approche naïve consisterait à choisir un certain temps t pour lequel nous calculerions les proportions de survivants au-delà de ce temps t dans les m groupes, puis de les comparer par un test du chi-deux. Cependant, cette approche serait totalement biaisée en cas de censure. De la même façon que pour l'estimateur de Kaplan-Meier, l'idée du test du Logrank est de construire une statistique de test basée sur les différents temps d'évènement survenant dans chaque groupe. La statistique de test du Logrank est basée sur la Table 2.1.

TABLE 2.1 – Table de contingence du nombre d'évènements et de l'effectif à risque par groupe au temps t_j

	Évènements	Survivants	À risque
Groupe 1	M_{1j}	$R_{1j} - M_{1j}$	R_{1j}
...
Groupe m	M_{mj}	$R_{mj} - M_{mj}$	R_{mj}
TOTAL	M_j	$R_j - M_j$	R_j

Soit $j \in \{1, \dots, k\}$. Sous \mathcal{H}_0 , le nombre d'évènements M_j au temps t_j est distribué proportionnellement entre les groupes. Ainsi, pour tout $l \in \{1, \dots, m\}$, le nombre d'évènements attendus au temps t_j dans le groupe l est donné par

$$M_j \times \frac{R_{lj}}{R_j},$$

et la statistique de test du Logrank, notée Z^{KM} , vaut

$$Z^{\text{KM}} = \sum_{l=1}^m \left[\frac{\left(\sum_{j=1}^k M_{lj} - \sum_{j=1}^k M_j \times \frac{R_{lj}}{R_j} \right)^2}{\sum_{j=1}^k M_j \times \frac{R_{lj}}{R_j}} \right].$$

Sous \mathcal{H}_0 , $Z^{\text{KM}} \stackrel{\mathcal{L}}{\sim} \chi^2(m-1)$ où $\stackrel{\mathcal{L}}{\sim}$ désigne la convergence en loi et $\chi^2(m-1)$ la loi du chi-deux à $(m-1)$ degrés de liberté. Ainsi, le test qui rejette \mathcal{H}_0 lorsque $Z^{\text{KM}} > C_{1-\alpha}^{m-1}$ est un test de niveau α , avec $C_{1-\alpha}^{m-1}$ le quantile d'ordre $(1-\alpha)$ de la loi $\chi^2(m-1)$.

2.3 Modèle à risques proportionnels : le modèle de Cox

2.3.1 Modèle de Cox

Le modèle de Cox a été développé par COX 1972. Son objectif est d'évaluer l'effet de différentes covariables sur la survie des individus. Il s'applique sur des données de survie avec censure à droite, et troncature à gauche.

Pour appliquer ce modèle, il faut veiller à ce que certaines hypothèses soient vérifiées :

- l'indépendance entre les mesures,
- la proportionnalité des risques, c'est à dire que l'effet des variables explicatives est supposé constant au cours du temps,
- la log-linéarité des variables explicatives.

En pratique, nous supposons que les hypothèses d'indépendance des mesures et de log-linéarité des variables sont vérifiées. Afin de vérifier l'hypothèse de proportionnalité des risques, une première méthode graphique consiste à vérifier sur le graphe du même type que la Figure 2.4 si les différentes courbes de survie ne se croisent pas.

Une seconde méthode est basée sur les résidus de Schoenfeld (SCHOENFELD 1982). Le test statistique associé calcule la corrélation entre ces résidus et une fonction du temps basée sur l'estimateur de Kaplan-Meier, permettant de rejeter ou non l'hypothèse de proportionnalité des risques.

De plus, lorsque l'hypothèse de proportionnalité des risques est respectée, la représentation graphique des résidus de Schoenfeld montre une répartition totalement aléatoire et l'évolution moyenne dans le temps est une droite horizontale. Si ce n'est pas le cas, les résidus sont d'abord négatifs puis positifs, et l'évolution moyenne sera une courbe croissante ou décroissante.

2.3.1.1 Définition du modèle

Le modèle de Cox s'écrit de la manière suivante :

$$\lambda(t) = \lambda_0(t) \exp(X^\top \beta)$$

avec

- $\lambda_0(t)$ la fonction de risque baseline associée au temps t (probabilité de connaître l'évènement pour les individus dont les caractéristiques sont nulles à t) telle que

$$\lambda_0(t) = \lambda(t|X_1 = 0, \dots, X_p = 0),$$

- $X = (X_1, \dots, X_p)$ l'ensemble des p covariables invariables dans le temps,
- β le vecteur de dimension p contenant les coefficients de régression.

On distingue le modèle de Cox paramétrique du modèle de Cox semi-paramétrique : soit la forme de λ_0 est connue (loi de Weibull ou Gamma par exemple), soit aucune hypothèse n'est faite sur la distribution de λ_0 . Pour toute la suite de cette section, nous nous plaçons dans le modèle de Cox semi-paramétrique pour lequel seuls les coefficients de β sont à estimer.

2.3.1.2 Estimation du paramètre

Les coefficients du paramètre de régression β sont estimés par la méthode de maximisation de la log-vraisemblance partielle. Nous utilisons ce type de log-vraisemblance car seuls les individus ayant connu l'évènement durant l'étude sont pris en considération.

Soient $t_1 < t_2 < \dots < t_k$ les temps à l'évènement observés supposés tous distincts. Autrement dit, nous supposons qu'un seul évènement arrive au temps t_i , $i \in \{1, \dots, k\}$, parmi tous les individus à risque au temps t_i . Soit R_i l'ensemble des individus à risque juste avant t_i . Nous notons $(1), \dots, (k)$ les indices des individus ayant connu un évènement respectivement en t_1, \dots, t_k . Pour tout $i \in \{1, \dots, k\}$, $p_i(t_i)$ désigne la probabilité que l'individu i connaisse l'évènement à t_i et dépend uniquement du paramètre de régression β . Elle vaut

$$p_i(t_i) = \frac{\lambda_0(t_i) \exp(X_{(i)}^\top \beta)}{\sum_{l \in R_i} \lambda_0(t_i) \exp(X_l^\top \beta)} = \frac{\exp(X_{(i)}^\top \beta)}{\sum_{l \in R_i} \exp(X_l^\top \beta)}.$$

La vraisemblance partielle est définie comme le produit de ces probabilités

$$L_{\text{cox}}(\beta) = \prod_{i=1}^k \frac{\exp(X_{(i)}^\top \beta)}{\sum_{l \in R_i} \exp(X_l^\top \beta)}.$$

En prenant le logarithme, la log-vraisemblance partielle est

$$l_{\text{cox}}(\beta) = \ln \left[\prod_{i=1}^k \frac{\exp(X_{(i)}^\top \beta)}{\sum_{l \in R_i} \exp(X_l^\top \beta)} \right] = \sum_{i=1}^k \left[X_{(i)}^\top \beta - \ln \left(\sum_{l \in R_i} \exp(X_l^\top \beta) \right) \right].$$

Pour tout $j \in \{1, \dots, p\}$, les estimateurs $\hat{\beta}_j$ des coefficients de régression β_j sont donc ceux qui maximisent $l_{\text{cox}}(\beta)$, soient les solutions de

$$\frac{\partial l_{\text{cox}}(\beta)}{\partial \beta_j} = 0.$$

2.3.1.3 Interprétation des coefficients du paramètre

Considérons un modèle de Cox à une seule variable explicative X_1 . Supposons tout d'abord X_1 binaire, où $X_1 = 1$ si l'individu est un homme et $X_1 = 0$ si l'individu est une femme. Soient deux individus X^* et X^+ , où $X_1^* = 1$ et $X_1^+ = 0$.

Le *Hazard Ratio* (HR) dans ce cas est :

$$HR = \frac{\lambda(t | X^*)}{\lambda(t | X^+)} = \frac{\lambda_0(t) \exp(X_1^* \beta_1)}{\lambda_0(t) \exp(X_1^+ \beta_1)} = \exp(\beta_1).$$

On interprète cette quantité de la façon suivante : le risque instantané de connaître l'évènement pour les hommes correspond à $\exp(\beta_1)$ fois celle des femmes.

Supposons désormais X_1 continue. Alors, $\exp(\beta_1)$ correspondrait au HR lors d'une augmentation d'une unité de la variable X_1 . Il est constant pour cette même augmentation qu'importe la valeur de X_1 : c'est l'hypothèse de log-linéarité qui est vérifiée. Elle peut cependant être mise à mal. En effet, si nous étudions une maladie touchant principalement les personnes âgées et que l'âge était considéré en covariable, le modèle supposerait que le HR serait le même pour l'augmentation d'une unité à 20 ans qu'à 80 ans.

De plus, il arrive aussi que l'évolution de variables explicatives au cours du suivi (comme le traitement d'un individu) peut avoir un impact sur la survenue de l'évènement d'intérêt. Il est donc important de suivre les individus afin d'observer ces potentiels changements, et pouvoir ainsi expliquer de manière fiable le temps à l'évènement. Dans ce cas, une des extensions du modèle classique de Cox est le modèle de Cox à variables dépendantes du temps présenté dans la Section 2.3.2 suivante.

2.3.2 Modèle de Cox dépendant du temps

Ce modèle permet de prendre en considération des covariables qui dépendent du temps. C'est une extension du modèle de Cox classique et nécessite la validité de certaines hypothèses (VAN HOUWELINGEN et PUTTER 2011). Les hypothèses d'indépendance des mesures intra et inter-individu, et de proportionnalité des risques doivent être vérifiées. De plus, il faut veiller à connaître la valeur de toutes les covariables à chaque temps t observé.

Le modèle de Cox dépendant du temps peut s'écrire de deux façons :

$$\lambda(t) = \lambda_0(t) \exp(X^\top(t)\beta(t)) \quad (2.3)$$

$$\text{ou } \lambda(t) = \lambda_0(t) \exp(X^\top(t)\beta(t) + Z^\top\gamma) \quad (2.4)$$

avec

- $\lambda_0(t)$ la fonction de risque baseline,
- $\beta(t)$ le vecteur de dimension p contenant les coefficients de régression dépendants du temps, associé au vecteur $X(t) = (X_1(t), \dots, X_p(t))$ des p covariables elles aussi dépendantes du temps,
- γ le vecteur de dimension q contenant les coefficients de régression fixes, associé à $Z = (Z_1, \dots, Z_q)$ l'ensemble des q covariables non-dépendantes du temps.

L'équation (2.3) est utilisée lorsque toutes les variables sont dépendantes du temps, tandis que l'équation (2.4) sert lorsque certaines variables sont invariantes dans le temps.

L'hypothèse la plus forte considérée par le modèle de Cox dépendant du temps est l'hypothèse d'indépendance entre les différentes mesures intra-individu. Par exemple, si nous considérons un groupe d'individus malades qui changerait de traitement au cours de l'étude à un moment t , l'indépendance des données pour chaque sujet entre le début de l'étude et après ce changement ne serait plus observée. Nous pouvons nous attendre à un changement des caractéristiques des patients, et donc à un changement de trajectoire de la variable expliquée. Une modélisation alternative permettant de relaxer cette hypothèse est présentée en Section 2.3.3 suivante.

2.3.3 Modélisation landmark : un raffinement du modèle de Cox

La modélisation landmark a été développée par VAN HOUWELINGEN 2007. L'idée de cette modélisation est de faire des analyses de survie sur des intervalles de temps plus courts que la durée de l'étude afin de prendre en compte d'éventuels changements lors de cette dernière. En reprenant l'exemple d'un groupe d'individus qui changerait de traitement au cours de l'observation, la modélisation landmark permet dans ce cas d'établir deux modèles différents sur deux fenêtres de temps différentes. Le premier intervalle de temps commencerait au début de l'étude et finirait au moment du changement ; puis un second irait du moment où le nouveau traitement a été introduit jusqu'à la fin de l'étude.

Plus généralement, décrivons le déroulement de la modélisation landmark. Il faut d'abord commencer par découper notre base en plusieurs fenêtres d'observation comme sur la Figure 2.5.



FIGURE 2.5 – Modélisation landmark

Le temps de début de la fenêtre d'observation $s \in \mathbb{R}^+$ est appelé temps landmark, et $h \in \mathbb{R}^{+*}$ l'horizon de prédiction. Par la suite, nous construisons l'échantillon landmark associé à chaque fenêtre d'observation en ne sélectionnant que les individus à risque au début de celle-ci. Par exemple, sur la fenêtre d'observation de la Figure 2.5, nous ne sélectionnons que les individus à risque au temps s . Nous définissons alors $X(s)$ l'ensemble des covariables au temps s , et nous considérons sur cette fenêtre d'observation le modèle de Cox avec variables dépendantes du temps suivant pour tout $t \in [s; s + h]$

$$\lambda(t|X(s), s) = \lambda_0(t|s) \exp(X^\top(s)\beta(s)).$$

Pour ce modèle, notons que l'hypothèse de proportionnalité des risques entre les différents échantillons landmark n'est plus à vérifier. Enfin, après avoir obtenu les estimations $\hat{\beta}$ et $\hat{\lambda}_0$ du vecteur des coefficients de régression β et de la fonction de risque baseline λ_0 , la probabilité de survie au-delà de $s + h$ s'obtient par

$$\mathbb{P}(T > s + h | T > s, X(s)) = \exp \left[- \exp(X^\top(s)\hat{\beta}(s)) \hat{\lambda}_0(s + h | s) \right].$$

Ces prédictions dépendent uniquement des caractéristiques connues au début de la fenêtre d’observation. Elles peuvent changer et être mises à jour au fil du temps. L’évolution de la survie au temps $s + h$ pour un individu dépend donc seulement des caractéristiques au temps s . Elles sont très utiles puisqu’elles permettent de connaître la probabilité d’évènement pour un individu à partir de ses propres caractéristiques.

Cependant, il arrive parfois qu’un évènement annexe puisse empêcher la survenue de l’évènement étudié. Si l’évènement d’intérêt est par exemple l’apparition d’une maladie, le décès sera considéré comme un évènement concurrent puisque tout individu décédé ne pourra plus développer cette maladie. La Section 2.4 présente une introduction à la théorie des risques concurrents.

2.4 Risques concurrents

Dans un cadre classique d’analyse de survie (voir Section 2.2), nous considérons que les individus sont à risque de subir un unique évènement. Dans le cadre des risques concurrents, les individus sont susceptibles d’en subir plusieurs. Nous pouvons représenter les modèles à risques concurrents par la Figure 2.6.

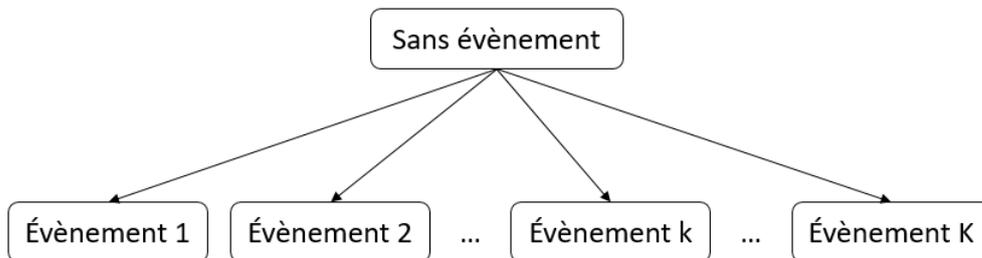


FIGURE 2.6 – Diagramme d’un modèle à K risques concurrents, $k \in \{1, \dots, K\}$, $K \geq 2$

Cela s’apparente à un modèle multi-états avec un stade initial « Sans évènement » et les différents types d’évènements terminaux existants. Les évènements concurrents sont mutuellement exclusifs : la survenue d’un type (d’une cause) d’évènement empêche l’apparition d’un autre évènement.

Exemple 7. *En s’intéressant à une population de patients atteints de la NAFLD (état « Sans évènement » de la Figure 2.6), nous pouvons considérer que les patients peuvent*

décéder de trois causes ($K = 3$, $k \in \{1, 2, 3\}$) : un cancer du foie ($k = 1$), une décompensation de cirrhose ($k = 2$) ou une autre cause ($k = 3$). Dans ce cas, les patients peuvent subir trois risques concurrents et un patient décédé d'un cancer du foie ne pourra pas décéder d'une décompensation de cirrhose (événements exclusifs).

2.4.1 Fonctions associées aux distributions de survie

Dans cette partie, les fonctions associées aux distributions de survie préalablement définies dans la Section 2.2.1 et adaptées au cadre des risques concurrents sont présentées.

2.4.1.1 Définitions

Les modèles à risques concurrents sont définis par des fonctions de risque cause-spécifique. Soient T le temps entre l'inclusion et l'occurrence du premier événement (date de dernière nouvelle sinon) et $K \geq 2$ le nombre d'événements concurrents. Dans ce contexte, notons $\lambda_k(t)$ la fonction de risque instantané au temps t spécifique à l'événement k , $k \in \{1, \dots, K\}$. Elle est définie par

$$\lambda_k(t) = \lim_{\Delta t \rightarrow 0} \frac{\mathbb{P}(t \leq T \leq t + \Delta t, \eta = k \mid T \geq t)}{\Delta t}. \quad (2.5)$$

Définie en Section 2.2.1, la fonction de risque globale $\lambda(t)$ ne tient pas compte du type de l'événement. Elle est égale à la somme des K fonctions de risque cause-spécifique et est définie par

$$\lambda(t) = \lim_{\Delta t \rightarrow 0} \frac{\mathbb{P}(t \leq T \leq t + \Delta t \mid T \geq t)}{\Delta t} = \sum_{k=1}^K \lambda_k(t).$$

Le terme $\lambda(t) \Delta t$ (respectivement $\lambda_k(t) \Delta t$) correspond à la probabilité pour un individu de subir n'importe quel événement (respectivement l'événement k) entre t et $(t + \Delta t)$ sachant qu'il est encore à risque au temps t . En partant de l'équation (2.5), la fonction de survie au temps t , interprétée comme la probabilité de ne subir aucun événement avant t , est

$$S(t) = \exp \left(- \sum_{k=1}^K \int_0^t \lambda_k(s) ds \right) = \exp \left(- \sum_{k=1}^K \Lambda_k(t) \right)$$

avec $\Lambda_k(t) = \int_0^t \lambda_k(s) ds$ la fonction de risque cumulé cause-spécifique. Enfin, appelée fonction d'incidence cumulée de l'événement k , la probabilité de subir l'événement k avant

le temps t est définie par

$$F_k(t) = \mathbb{P}(T \leq t, \eta = k) = \int_0^t \lambda_k(u) S(u) du. \quad (2.6)$$

2.4.1.2 Estimations

De la même façon que pour l'estimateur de Kaplan-Meier de la fonction de survie $S(t)$ au temps t , nous définissons $0 < t_1 < t_2 < \dots < t_f$ la discrétisation de l'intervalle de temps $[0; t_f]$ correspondant aux durées observées jusqu'à chaque survenue de n'importe quel évènement k , $k \in \{1, \dots, K\}$. Soient R_i le nombre d'individus encore à risque au temps t_i , $i \in \{1, \dots, f\}$, M_{ki} le nombre d'individus ayant subi l'évènement k à t_i et M_i le nombre d'individus ayant subi n'importe quel évènement k à t_i . Nous avons alors

$$M_i = \sum_{k=1}^K M_{ki}.$$

Rappelons que l'estimateur de Kaplan-Meier (voir Section 2.4) pour la fonction de survie sans considérer de risque concurrent est

$$\hat{S}(t) = \prod_{i:t_i \leq t} \left(1 - \frac{M_i}{R_i}\right). \quad (2.7)$$

La discrétisation de l'équation (2.5) du risque cause-spécifique est donnée par

$$\lambda_k(t_i) = \mathbb{P}(T = t_i, \eta = k \mid T > t_{i-1}),$$

et estimée par la proportion d'individus ayant subi l'évènement k à t_i

$$\hat{\lambda}_k = \frac{M_{ki}}{R_i}.$$

Dans le contexte des risques concurrents, l'estimation de $S(t)$ est donnée à l'aide de l'équation (2.7) par

$$\hat{S}(t) = \prod_{i:t_i \leq t} \left(1 - \sum_{k=1}^K \hat{\lambda}_k(t_i)\right);$$

et celle de la fonction d'incidence cumulée de l'évènement k est estimée grâce à l'équation (2.6) par

$$\hat{F}_k(t) = \sum_{i:t_i \leq t} \hat{\lambda}_k(t_i) \hat{S}(t_{i-1}).$$

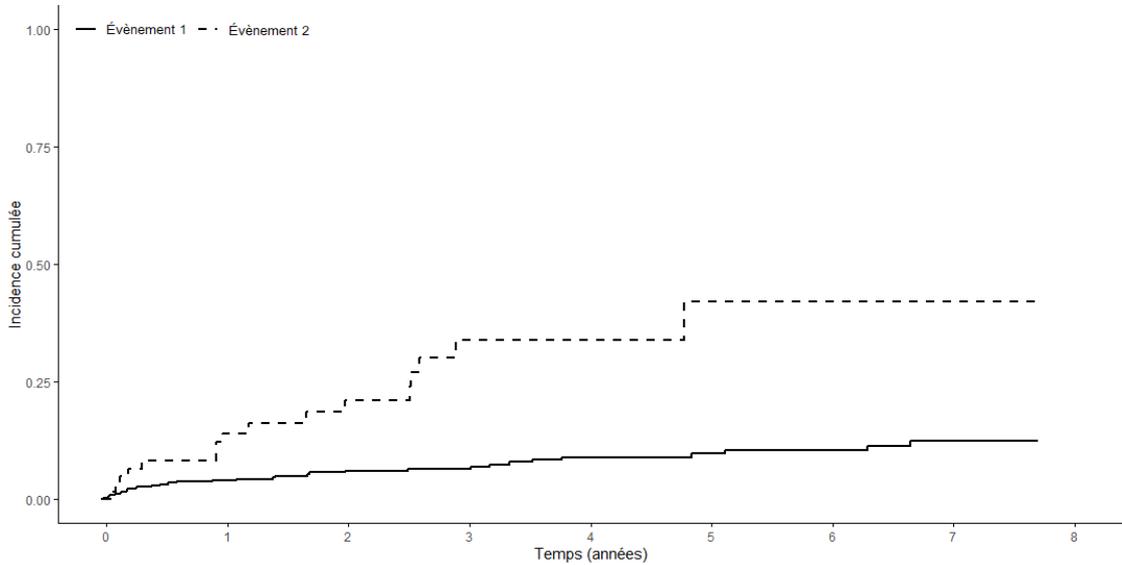


FIGURE 2.7 – Courbes d'incidence cumulée pour deux événements concurrents observés

Exemple 8. Dans un cas où deux événements concurrents ($k = 1, 2$) sont observés dans une population, la Figure 2.7 présente les deux courbes d'incidence obtenues (équation (2.6)) pour chacun des deux événements.

2.4.2 Modèles de régression

Différentes méthodes sont utilisées en statistiques afin d'estimer les effets des variables dans le cadre des risques concurrents. Trois d'entre elles sont proposées dans cette section.

2.4.2.1 Régression sur des risques cause-spécifique

Le modèle à risques proportionnels sur un risque cause-spécifique semble être le plus approprié lorsque l'on souhaite étudier les effets des covariables sur un événement $k \in \{1, \dots, K\}$ d'intérêt. Le modèle s'écrit de la façon suivante :

$$\lambda_k(t | X) = \lambda_{k,0}(t) \exp(X^\top \beta_k)$$

avec

- $\lambda_{k,0}(t)$ la fonction de risque baseline cause-spécifique associé à l'évènement k ,
- $X = (X_1, \dots, X_p)$ l'ensemble des p covariables observées,
- $\beta_k = (\beta_{k,1}, \beta_{k,2}, \dots, \beta_{k,p})^\top$ le vecteur des effets des covariables sur l'évènement k .

Les individus ayant subi des évènements différents que celui d'intérêt k sont considérés comme censurés : l'inférence de ce modèle est donc la même que pour un modèle sans évènement concurrent.

2.4.2.2 Régression sur plusieurs types d'évènements

Le modèle de régression sur plusieurs types d'évènements est une extension du modèle Cox. Il permet de tester si l'effet d'une covariable diffère selon les différents évènements observés. Pour tout $\beta_k^{\text{int}} \in \mathbb{R}$, $k \in \{1, \dots, K\}$, le modèle est défini par

$$\lambda_k(t | X) = \lambda_1(t) \exp(\beta_k^{\text{int}} + X^\top \beta_k + X^\top \gamma)$$

avec

- $\lambda_k(t) = \lambda_1(t) \exp(\beta_k^{\text{int}})$ la fonction de risque baseline de l'évènement $k \neq 1$, et $\lambda_1(t)$ la fonction de risque pour l'évènement $k = 1$ (covariables nulles) avec $\beta_1^{\text{int}} = 0$,
- $\beta_k = (\beta_{k,1}, \beta_{k,2}, \dots, \beta_{k,p})^\top$ le vecteur des effets des covariables sur l'évènement k avec $\beta_1 = (0, \dots, 0)$,
- $X = (X_1, \dots, X_p)$ l'ensemble des p covariables observées,
- $\gamma = (\gamma_1, \gamma_2, \dots, \gamma_p)^\top$ le vecteur des effets des covariables commun à tous les évènements.

Remarque 2. Comme $\lambda_k(t) = \lambda_1(t) \exp(\beta_k^{\text{int}})$ désigne la fonction de risque baseline de l'évènement $k \neq 1$, l'hypothèse de proportionnalité des risques entre les différents évènements est imposée.

Le terme $\beta_k^\top X$ permet de tester si l'effet d'une covariable est la même pour deux évènements. Pour tester par exemple l'effet de la cinquième covariable X_5 , les risques relatifs pour l'augmentation d'une unité sur X_5 sont donnés par $\exp(\gamma_5)$ pour l'évènement $k = 1$ et $\exp(\gamma_5 + \beta_{k,5})$ pour l'évènement $k \neq 1$. Ainsi, tester si l'effet de X_5 est le même pour deux évènements 1 et $k \neq 1$ revient à tester si le coefficient $\beta_{k,5}$ est égal à zéro.

2.4.2.3 Régression sur les fonctions d'incidence cumulée

Une autre forme de modèle semi-paramétrique à risques proportionnels basé sur la fonction d'incidence cumulée a été développée par FINE et GRAY 1999. Ils proposent de

modéliser l'effet des covariables directement sur les fonctions d'incidence cumulée au lieu de combiner les différents risques cause-spécifique comme précédemment. Pour cela, une fonction de risque associée à la fonction d'incidence cumulée, aussi appelée fonction de risque de sous-répartition, est définie :

$$\bar{\lambda}_k(t) = -\frac{d \log(1 - F_k(t))}{dt}.$$

La fonction de risque de sous-répartition $\bar{\lambda}_k(t)$ est différente de celle du risque cause-spécifique $\lambda_k(t)$. Dans le cas de $\lambda_k(t)$, le nombre d'individus à risque diminue à chaque fois qu'un évènement $k \in \{1, \dots, K\}$ survient. Tandis que pour $\bar{\lambda}_k(t)$, les individus ayant subi un évènement différent de l'évènement k observé sont toujours comptabilisés dans l'effectif à risque. Par exemple, si des individus sont censurés lors du suivi sans connaître d'évènement, alors ils restent dans l'effectif à risque jusqu'à leur temps de censure. En revanche, si des individus ne sont pas censurés et subissent un évènement différent de k , ils restent dans l'effectif à risque pour un temps « infini ». Ces individus se voient alors attribuer un temps de censure plus grand que tous les temps d'évènement pour rendre l'analyse standard.

Ce modèle fait l'hypothèse de risques proportionnels sur les risques de sous-répartition :

$$\bar{\lambda}_k(t | X = x) = \bar{\lambda}_{k,0}(t) \exp(X^\top \beta_k) \quad (2.8)$$

avec

- $\bar{\lambda}_{k,0}(t)$ la fonction de risque baseline de l'évènement k ,
- $\beta_k = (\beta_{k,1}, \beta_{k,2}, \dots, \beta_{k,p})^\top$ le vecteur des effets des covariables sur l'évènement k ,
- $X = (X_1, \dots, X_p)$ l'ensemble des p covariables observées,
- $\exp(\beta_k)$ le risque relatif associé à la covariable x .

L'estimation des paramètres se fait sur le même principe de vraisemblance partielle utilisé lors du modèle de Cox dans la Section 2.3.1.2. De plus, il existe un test développé par GRAY 1988 permettant de comparer les courbes d'incidence obtenues pour chaque évènement concurrent.

Les notions développées dans la suite de ce chapitre existent dans le cadre des risques concurrents. Cependant, par souci de simplification des notations, elles sont définies uniquement dans le cadre classique avec un seul évènement d'intérêt.

2.5 Modèle conjoint

Les modèles conjoints (HENDERSON, DIGGLE et DOBSON 2000) s’appliquent sur des bases de données combinant des données longitudinales et des données de survie. Ils sont composés d’un modèle linéaire à effets mixtes pour les mesures répétées de la variable à expliquer, d’un modèle de survie pour le risque individuel de l’évènement et d’une structure latente commune inobservable dans la réalité afin de lier les deux processus. La Figure 2.8 présente synthétiquement la structure globale d’un modèle conjoint.

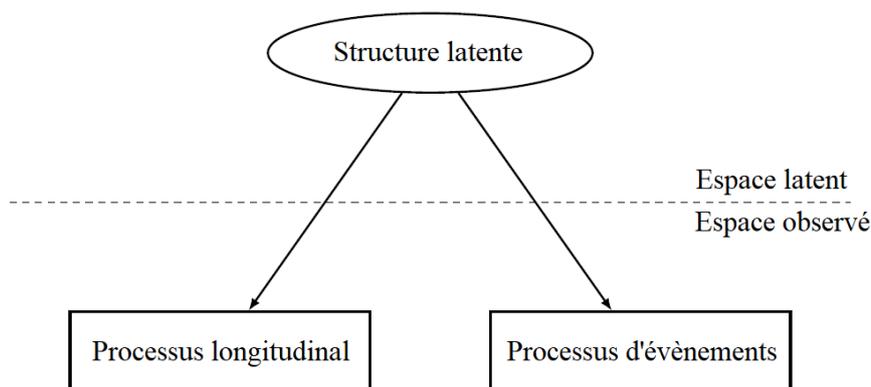


FIGURE 2.8 – Graphique de la structure d’un modèle conjoint (FERRER 2017, Fig. 2.4)

Les modèles conjoints permettent d’étudier l’évolution d’une variable durant le suivi d’un individu, mais aussi de comprendre le lien entre la variable étudiée et l’évènement d’intérêt. Ils rendent possible l’analyse simultanée des données longitudinales et des données de survie.

Il existe deux types de modèles conjoints. Le premier est celui des modèles conjoints à effets aléatoires partagés (TSIATIS et DAVIDIAN 2004; RIZOPOULOS, VERBEKE et MOLENBERGHS 2008). Dans ce cas, tous les individus ont des caractéristiques similaires. L’association entre le processus longitudinal et le processus d’évènement est alors effectuée grâce à une fonction d’effets aléatoires individuels. Elle permet notamment d’expliquer la variabilité des différentes trajectoires longitudinales entre les individus. Le second est celui des modèles conjoints à classes latentes (PROUST-LIMA et TAYLOR 2009). Ici, la population est divisée en plusieurs classes latentes : les individus semblent avoir des caractéristiques qui diffèrent et forment plusieurs groupes. L’échantillon de population contenu dans chaque classe est quant à lui supposé homogène. Ces classes sont caractérisées par une trajectoire moyenne du processus longitudinal et un risque d’évènement

qui sont propres à chacune d'entre elles. Nous développons par la suite ces deux types de modèle conjoint.

2.5.1 Modèle conjoint à effets aléatoires partagés

Le modèle à effets aléatoires partagés (RIZOPOULOS, VERBEKE et MOLENBERGHS 2008) associe deux sous-modèles : un modèle linéaire à effets mixtes pour les mesures répétées de la variable à expliquer (comme défini en Section 2.1.1), d'un modèle de survie pour le risque individuel de l'évènement d'intérêt (comme défini en Section 2.3.1), et d'une fonction liant les deux modèles.

2.5.1.1 Définition du modèle conjoint à effets aléatoires partagés

En conservant les mêmes notations que dans les Sections 2.1.1 et 2.3.1, le modèle conjoint à effets aléatoires partagés se définit formellement, pour tout $i \in \{1, \dots, N\}$, pour tout $j \in \{1, \dots, n_i\}$, par

$$\begin{cases} Y_i^{(j)} &= X_{i_j}^{\text{mxt}\top} \beta + Z_{i_j}^\top b_i + \epsilon_i^{(j)} \\ \lambda_i(t) &= \lambda_0(t) \exp(X_i^{\text{surv}\top} \gamma + g_i(b_i, t)^\top \eta) \end{cases}$$

où

- $Y_i^{(j)}$ est la variable aléatoire réponse associée à la j -ème mesure de l'individu i ,
- $\beta \in \mathbb{R}^{p+1}$, $p \in \mathbb{N}^*$, est le vecteur des effets fixes associé aux covariables $X_{i_j}^{\text{mxt}}$,
- $b_i \in \mathbb{R}^q$, $q \in \mathbb{N}^*$, $q \leq p$, est le vecteur des effets aléatoires associé au vecteur de covariables Z_{i_j} , où Z_{i_j} est un sous-vecteur de $X_{i_j}^{\text{mxt}}$. Les b_i sont des vecteurs gaussiens i.i.d. centrés de matrice de variance-covariance B ,
- $\epsilon_i^{(j)}$ est un résidu gaussien centré de variance $\sigma^2 > 0$, avec $\epsilon_i^{(j)} \perp\!\!\!\perp \epsilon_i^{(j')}$, pour tout $j \neq j'$, $j, j' \in \{1, \dots, n_i\}$, et $\epsilon_i^{(\cdot)} \perp\!\!\!\perp b_i$, $i \in \{1, \dots, N\}$,
- $\lambda_0(t)$ est la fonction de risque baseline associée au temps t (paramétrique ou non),
- $\gamma \in \mathbb{R}^r$, $r \in \mathbb{N}^*$, est le vecteur contenant les coefficients de régression pour le modèle de survie associé au vecteur de covariables relatif à la survie X_i^{surv} ,
- η est le coefficient de régression associé à la fonction de lien $g_i(b_i, t)$.

La fonction $g_i(b_i, t)$ qui associe les deux sous-modèles entre eux peut être uni ou multi-dimensionnelles. Elle est donc définie de différentes manières en fonction du contexte et de l'objectif de l'étude statistique. Deux formes sont classiquement rencontrées dans la littérature. D'une part, $g_i(b_i, t) = Y_i(t)$, avec $Y_i(t) = X_i^{\text{mxt}\top} \beta + Z_i^\top b_i$. La variable réponse Y_i dépend du temps à travers les covariables X_i^{mxt} et Z_i . Dans ce cas, le risque instantané de l'évènement au temps t dépend uniquement des covariables explicatives associées au modèle de survie et de la valeur de la variable à expliquer à t sans tenir compte de l'erreur de mesure. C'est l'un des modèles les plus utilisés. Une définition plus flexible de g_i est donnée par $g_i(b_i, t) = \left(Y_i(t); \frac{\partial Y_i(t)}{\partial t} \right)$. Dans ce cas, nous considérons que le risque instantané de l'évènement à t dépend de la vraie valeur de Y_i à t , mais aussi de la pente de celle-ci.

2.5.1.2 Estimation des paramètres

Les paramètres à estimer sont : les paramètres de régression β , γ et η , les paramètres de variance B et σ appartenant à la partie du modèle mixte, ainsi que la fonction baseline λ_0 pour la partie survie. La modélisation conjointe prend en compte la corrélation qu'il pourrait y avoir entre le processus longitudinal et le processus de survie, et corrige ce biais en se basant sur des méthodes utilisant la vraisemblance complète jointe. Sa maximisation permet l'estimation des différents paramètres.

Soit $\theta_{\text{JM}} = (\beta, \gamma, \eta, B, \sigma, \lambda_0)$ l'ensemble des paramètres à estimer. Comme Y et b suivent une loi normale, la densité conditionnellement aux effets aléatoires, notée $f_{Y|b}$, vaut

$$f_{Y|b}(Y_i|b_i) = \frac{1}{(2\pi\sigma^2)^{\frac{n_i}{2}}} \exp\left(\frac{-\|Y_i - X_i^{\text{mxt}\top} \beta - Z_i^\top b_i\|^2}{2\sigma^2}\right),$$

où $\|\cdot\|^2$ désigne la norme euclidienne, et la densité des effets aléatoires, notée f_b , s'écrit

$$f_b(b_i) = \frac{1}{(2\pi)^{\frac{q}{2}} |B|^{\frac{1}{2}}} \exp\left(\frac{-b_i^\top B^{-1} b_i}{2}\right).$$

Par ailleurs, la densité associée au modèle de survie conditionnellement aux effets aléatoires, notée $f_{T|b}$ est donnée par

$$f_{T|b}(T_i|b) = \lambda_i(T_i|b) \exp\left(-\int_0^{T_i} \lambda_i(s|b) ds\right).$$

Les données longitudinales étant indépendantes des données de survie conditionnellement

aux effets aléatoires partagés, la log-vraisemblance peut donc s'écrire de la façon suivante :

$$l_{\text{JM}}(\theta_{\text{JM}}) = \sum_{i=1}^N \ln \left(\int f_{Y|b}(Y_i|b_i) f_T(T_i, \delta_i|b) f_b(b_i) db \right).$$

De la même façon que pour les modèles linéaires mixtes à classes latentes, la Section 2.5.2 suivante présente les modèles conjoints à classes latentes.

2.5.2 Modèle conjoint à classes latentes

Contrairement aux modèles conjoints à effets aléatoires partagés où il n'y a qu'une seule association de processus longitudinal et de survie à modéliser, les modèles conjoints à classes latentes (PROUST-LIMA et TAYLOR 2009) possèdent autant de processus longitudinaux et de survie à estimer que de classes latentes. Chaque classe possède donc un modèle particulier de variation de Y et un risque spécifique à l'évènement.

2.5.2.1 Définition du modèle conjoint à classes latentes

Les modèles conjoints à classes latentes combinent trois sous-modèles : un modèle linéaire à effets mixtes pour expliquer la trajectoire de la variable à expliquer, un modèle de survie pour étudier les évènements, et un modèle logistique multinomial pour définir les probabilités d'appartenance à une classe. En reprenant les notations de la Section 2.1.3.1, le modèle conjoint à classes latentes se définit par

$$\begin{cases} \pi_{ig} & = \mathbb{P}(c_i = g | X_i^{\text{mlt}}) = \frac{\exp(\xi_{0g} + X_i^{\text{mlt}\top} \xi_{1g})}{\sum_{l=1}^G \exp(\xi_{0l} + X_i^{\text{mlt}\top} \xi_{1l})}, & \xi_{0G} = 0, \quad \xi_{1G} = 0 \\ Y_i^{(j)} & = X_{i_j}^{\text{mxt}\top} \beta_g + Z_{i_j}^\top b_i + \epsilon_i^{(j)} \\ \lambda_i(t | c_i = g) & = \lambda_0(t; \zeta_g) \exp(X_i^{\text{surv}\top} \gamma_g) \end{cases}$$

avec

- π_{ig} la probabilité que l'individu i appartienne à la classe g selon le modèle logistique multinomial expliqué par le vecteur de covariables X_i^{mlt} . La classe G est fixée comme étant la classe de référence, d'où $\xi_{0G} = 0$ et $\xi_{1G} = 0$, afin de s'assurer de l'identifiabilité du modèle,

- β_g le vecteur des effets fixes associé à la classe g et au vecteur de covariables $X_{i_j}^{\text{mixt}}$,
- b_i le vecteur des effets aléatoires associé au vecteur de covariables Z_{i_j} , tel que $b_i \sim \mathcal{N}(\mu_g, B_g)$,
- $\epsilon_i = (\epsilon_i^{(j)})_{1 \leq j \leq n_i} \sim \mathcal{N}(0, \Sigma_i)$ tel que $\epsilon_i^{(j)}$ est indépendant de $\epsilon_i^{(j')}$, pour tout $j \neq j'$, $j, j' \in \{1, \dots, n_i\}$,
- $\lambda_0(t; \zeta_g)$ la fonction de risque baseline associée au temps t pour la classe g (paramétrique ou non),
- $\gamma_g \in \mathbb{R}^r$, $r \in \mathbb{N}^*$, le vecteur contenant les coefficients de régression pour le modèle de survie associé à la classe g et au vecteur de covariables X_i^{surv} .

Pour la partie concernant la survie, nous rappelons que la fonction $\lambda_i(t|c_i = g)$ est spécifique à chaque classe.

2.5.2.2 Estimation des paramètres

L'estimation des paramètres se fait par la méthode du maximum de vraisemblance. Nous pouvons décomposer la log-vraisemblance du modèle grâce à l'hypothèse d'indépendance conditionnelle entre les différentes mesures répétées de Y et le temps à l'évènement T . Notons θ_{lcmmm} le vecteur contenant l'ensemble des paramètres à estimer contenus dans les trois sous-modèles. Pour G fixé, nous obtenons alors la log-vraisemblance suivante :

$$l_{\text{LCMM}}(\theta_{\text{lcmmm}}) = \sum_{i=1}^N \ln \left(\sum_{g=1}^G \pi_{ig} f_{Y|c}(Y_i|c_i = g) \lambda_i(T_i|c_i = g) S_i(T_i|c_i = g) \right)$$

avec,

- $f_{Y|c}(Y_i|c_i = g)$ la fonction de densité des mesures répétées de Y spécifique à la classe g ; elle est gaussienne de moyenne $Z_i \mu_g + X_i^{\text{mixt}} \beta_g$ et de matrice de variance-covariance $Z_i B_g Z_i^T + \Sigma_i$,
- et $S_i(T_i|c_i = g)$ la fonction de survie propre à la classe g .

Comme indiqué lors de la Section 2.5.1.2, le nombre de classes latentes G est sélectionné en retenant le modèle minimisant le critère du BIC. Afin de mesurer l'efficacité prédictive des modèles conjoints à classes latentes, une mesure dérivée de la théorie de l'information a été proposée et est définie dans la Section 2.5.2.3 suivante.

2.5.2.3 Expected Prognostic Observed Cross-Entropy

L'*Expected Prognostic Observed Cross-Entropy* (EPOCE) est utilisée pour évaluer la précision prédictive d'un modèle conjoint à classes latentes (COMMENGES, LIQUET et PROUST-LIMA 2012). Si elle est appliquée sur des données de validation externe, l'EPOCE est estimée par le *Mean Prognosis Observed Loss* (MPOL). Cependant, elle est aussi adaptée pour une validation interne : dans ce cas, c'est le *Cross-Validated Prognosis Observed Loss* (CVPOL) qui est calculé.

Soit $s \geq 0$ un temps de prédiction fixé. Soit $f_{\tilde{T}|\bar{Y}(s),T-s}$ la fonction de densité du temps à l'évènement censuré à droite $\tilde{T} = \min(T, C_d)$ en $(T - s)$ conditionnellement à $\bar{Y}(s)$ et dérivée des modèles conjoints, avec $\bar{Y}(s)$ l'ensemble des mesures répétées de la variable aléatoire réponse jusqu'au temps s . L'EPOCE est définie par

$$EPOCE = \mathbb{E} \left[-\ln \left(f_{\tilde{T}|\bar{Y}(s),T-s} \mid T \geq s \right) \right].$$

Dans le cas où l'EPOCE est évaluée sur des données d'apprentissage, elle peut être estimée par validation croisée leave-one-out. L'estimateur obtenu, noté CVPOL, est donné par

$$CVPOL(s) = -\frac{1}{n^{(s)}} \sum_{i=1}^{n^{(s)}} F_i(\hat{\theta}_{\text{lcmm}}, s) + N \text{Trace} \left(H^{-1} K_s \right) \quad (2.9)$$

avec

- N le nombre total d'individus, $n^{(s)}$ le nombre d'individus à risque au temps s ,
- F_i la contribution individuelle à la log-vraisemblance conditionnelle définie pour tout $i \in (1, \dots, n^{(s)})$ dans le cadre des modèles conjoints à classes latentes par

$$F_i(\hat{\theta}_{\text{lcmm}}, s) = \ln \left(\frac{\sum_{g=1}^G \pi_{ig} f(\bar{Y}_i(s) \mid c_i = g; \theta_{\text{lcmm}}) \lambda_i(\tilde{T}_i \mid c_i = g; \theta_{\text{lcmm}})^{\mathbb{1}_{\{T \leq C_d\}}} S_i(\tilde{T}_i \mid c_i = g; \theta_{\text{lcmm}})}{\sum_{g=1}^G \pi_{ig} f(\bar{Y}_i(s) \mid c_i = g; \theta_{\text{lcmm}}) S_i(s \mid c_i = g; \theta_{\text{lcmm}})} \right),$$

- H la matrice hessienne de la log-vraisemblance conjointe,
- $K_s = \frac{1}{n^{(s)}(N-1)} \sum_{i=1}^N \mathbb{1}_{\{\tilde{T}_i \geq s\}} \hat{v}_i(s) \hat{d}_i^T$ avec $\hat{v}_i(s)$ et \hat{d}_i les gradients des contributions individuelles respectivement à la log-vraisemblance conditionnelle à s en utilisant

seulement $\bar{Y}_i(s)$ et à la log-vraisemblance conjointe calculée en θ_{lcmmm} en utilisant le vecteur total de mesures répétées Y_i .

Dans le cas d’une validation externe, l’EPOCE est estimée par le MPOL défini par

$$MPOL(s) = -\frac{1}{n^{(s)}} \sum_{i=1}^{n^{(s)}} F_i(\hat{\theta}_{\text{lcmmm}}, s).$$

Il est défini uniquement par la première composante de la somme de l’équation (2.9) du CVPOL. En effet, le deuxième terme $\{\text{Trace}(H^{-1}K_s)\}$ est utilisé pour corriger le sur-optimisme provenant de l’utilisation de la validation croisée approximée dans le cas de données d’apprentissage. Finalement, les avantages de l’EPOCE pour mesurer la précision prédictive sont multiples. En plus de pouvoir être appliquée sur des données de validation interne et externe, aucune hypothèse n’est faite concernant la distribution de la censure. Enfin, un intervalle de confiance à 95% de la différence d’EPOCE entre deux modèles peut être calculé afin de comparer la différence de précision prédictive entre les deux.

En clinique, les statistiques peuvent être utilisées afin de comprendre des phénomènes et faire de la prédiction : pour cela, des modèles comme présentés précédemment sont appliqués. L’EPOCE est utilisée dans l’évaluation de l’efficacité prédictive de modèles conjoints à classes latentes. D’autres outils ont été développés pour permettre d’évaluer et de comparer différents tests cliniques ou autres modèles existants. Ils sont présentés dans la Section 2.6 suivante.

2.6 Critères de performance

Afin d’évaluer les performances diagnostiques ou pronostiques d’un marqueur ou de prédictions d’un modèle, des outils mesurant la discrimination et la calibration sont disponibles. Par exemple, un modèle de prédiction aura de bonnes capacités de discrimination si les individus ayant un risque prédit élevé (respectivement bas) sont effectivement plus à risque de subir l’évènement étudié (respectivement moins à risque). Par ailleurs, un modèle sera bien calibré si, pour tout pourcentage α , parmi les individus qui ont un risque prédit de $\alpha\%$, α individus sur 100 subissent l’évènement.

De nombreux critères de performance permettant de mesurer la discrimination et la calibration existent (HUANG et al. 2020 ; BLANCHE 2013). Dans cette section, il est

question d'en présenter les plus populaires. Afin de faciliter l'interprétation des outils présentés dans cette section, nous nous plaçons dans le contexte où l'évènement d'intérêt correspond à l'apparition d'une maladie.

2.6.1 Évaluation des capacités de discrimination d'un marqueur

2.6.1.1 Courbe ROC et AUC

La courbe *Receiver Operating Characteristic* (ROC) est fréquemment utilisée en médecine afin d'évaluer le potentiel d'un marqueur à discriminer les individus malades des non-malades au sein d'une population (BLANCHE, DARTIGUES et JACQMIN-GADDA 2013b).

Soient M un marqueur continu, $c \in \mathbb{R}$ un certain seuil, et D une variable binaire égale à 1 si le patient est malade, 0 sinon. L'objectif de la courbe ROC est d'étudier la capacité de discrimination du marqueur M sur l'ensemble des valeurs seuils possibles. Pour cela, il faut construire un test indiquant si le patient est malade ou non de la façon suivante : il est positif si $M \geq c$, et négatif si $M < c$. En pratique, l'idée est d'éviter autant que possible les erreurs de diagnostic et de limiter le nombre d'individus testés « Faux Positifs » et « Faux Négatifs ». La courbe ROC est obtenue en traçant le taux d'individus « Vrais Positifs » en fonction du taux d'individus « Faux Positifs ». Les deux principaux critères de qualité employés pour quantifier les performances d'un test diagnostique sont la sensibilité et la spécificité.

Sensibilité La sensibilité mesure la capacité d'un test à indiquer un résultat positif alors qu'une hypothèse est vérifiée. Elle est définie comme la probabilité de trouver tous les individus « Vrais positifs ». Elle dépend du seuil c choisi et s'écrit de la façon suivante :

$$\begin{aligned} Se(c) &= \mathbb{P}(\text{« Vrais Positifs »}) \\ &= \mathbb{P}(\text{« Test positif »} | \text{« Malade »}) \\ &= \mathbb{P}(M \geq c | D = 1). \end{aligned}$$

Autrement dit, la sensibilité est la probabilité que le marqueur M soit supérieur ou égal à un seuil c sachant que l'individu est malade. Elle s'interprète comme la probabilité que le test soit positif sachant que le patient est malade. Par exemple, une sensibilité de 0.78 indique qu'il y a 78% d'individus « Vrais Positifs ».

Spécificité La spécificité traduit la capacité d'un test à donner un résultat négatif quand une hypothèse n'est pas vérifiée. Elle est définie comme la probabilité de trouver tous les individus « Vrais Négatifs ». Elle dépend aussi du seuil c fixé et est définie par

$$\begin{aligned} Sp(c) &= \mathbb{P}(\text{« Vrais Négatifs »}) \\ &= \mathbb{P}(\text{« Test négatif »} | \text{« Pas malade »}) \\ &= \mathbb{P}(M < c | D = 0). \end{aligned}$$

En d'autres termes, la spécificité est la probabilité que le marqueur M soit strictement inférieur à un seuil c sachant que l'individu n'est pas malade. Elle s'interprète comme la probabilité que le test soit négatif sachant que le patient n'est pas malade. Une spécificité de 0.88 indique par exemple une probabilité de 88% d'individus « Vrais Négatifs ».

Courbe ROC Ainsi, la courbe ROC se définit comme l'ensemble des points tels que

$$\{(1 - Sp(c), Se(c)), c \in \mathbb{R}\},$$

où $Se(c)$ désigne la sensibilité et $Sp(c)$ la spécificité du test au seuil c . C'est une fonction croissante située dans le rectangle $[0; 1] \times [0; 1]$, valant 0 en 0, et 1 en 1. En pratique, le seuil c est souvent choisi de façon à réaliser le meilleur compromis entre sensibilité et spécificité : repérer le plus possible de patients « Vrais Positifs » en excluant un maximum de « Vrais Négatifs ».

Aire sous la courbe ROC Dans le but d'évaluer la capacité à discriminer les patients malades des non-malades d'un marqueur M , nous pouvons calculer l'aire sous la courbe ROC, aussi appelée AUC (*Area Under the ROC Curve*). Elle est comprise entre 0 et 1 et vaut

$$AUC = \int_0^1 Se \left[(1 - Sp)^{-1}(u) \right] du = \mathbb{P}(M_i \geq M_j | D_i = 1, D_j = 0).$$

C'est la probabilité que le marqueur d'un individu i soit supérieur ou égal à celui d'un individu j , sachant que i est malade, mais pas j (où i et j sont deux individus indépendants). En pratique clinique, si cette aire est inférieure à 0.20 ou supérieure à 0.80, nous considérerons que la capacité de discrimination du marqueur est plutôt bonne.

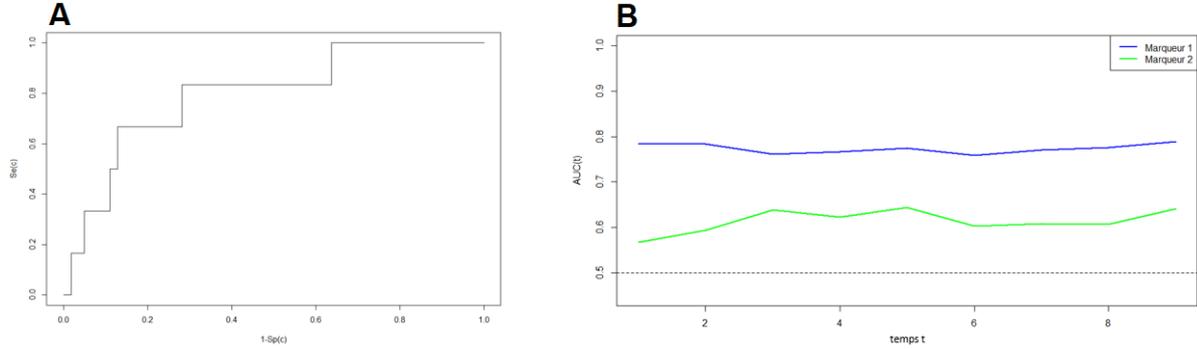


FIGURE 2.9 – Courbe ROC (A) et AUCs dépendants du temps (B)

Exemple 9. La partie A de la Figure 2.9 présente un exemple d'une courbe ROC où l'AUC est de 0.78. Le marqueur utilisé ne sera donc pas considéré comme satisfaisant pour discriminer les patients malades des non-malades.

AUC dépendant du temps Comme mentionné précédemment, l'objectif des courbes ROC est d'évaluer les performances d'un test de diagnostic. Un moyen d'évaluer celles d'un test pronostique est d'utiliser des courbes ROC dépendantes du temps (BLANCHE, DARTIGUES et JACQMIN-GADDA 2013a). Le but dans ce cas est d'évaluer la capacité d'un marqueur à détecter des individus qui pourraient développer une maladie dans le futur.

Soit M^0 un marqueur mesuré à l'inclusion et T la variable aléatoire représentant le moment d'apparition de l'évènement. Les termes de sensibilité $Se(c, t)$ et de spécificité $Sp(c, t)$ à l'instant t associés au seuil c sont définis par

$$Se(c, t) = \mathbb{P}(M^0 \geq c | T \leq t) \text{ et } Sp(c, t) = \mathbb{P}(M^0 < c | T > t).$$

La courbe ROC dépendante du temps est alors définie par l'ensemble des points

$$\{(1 - Sp(c, t), Se(c, t)), c \in \mathbb{R}\}.$$

Nous pouvons maintenant calculer l'AUC dépendant du temps associé à la mesure du marqueur à l'inclusion :

$$AUC(t, M_0) = \int_0^1 Se \left[(1 - Sp)^{-1}(u) \right] du = \mathbb{P}(M_i \geq M_j | T_i \leq t, T_j > t).$$

Autrement dit, l'AUC dépendant du temps représente la probabilité que la mesure du

marqueur de l’individu i soit supérieure ou égale à celle de l’individu j , sachant que l’individu i a connu l’évènement au temps t , mais pas l’individu j (où i et j sont deux individus indépendants).

Exemple 10. *La partie **B** de la Figure 2.9 présente l’évolution de l’AUC dépendant du temps pour deux marqueurs. Ici, le marqueur 1 (courbe bleue) semble détenir un pouvoir pronostique plus important que le marqueur 2 (courbe verte). En effet, l’AUC associé au premier marqueur reste toujours supérieur à celui associé au deuxième au cours du temps.*

2.6.1.2 C-index

Le C-index de Harrell (HARRELL JR, LEE et MARK 1996) est une extension directe de l’AUC du temps qui permet également de discriminer les individus malades des autres. Il est décrit comme une mesure de concordance entre un marqueur et le temps à l’évènement. Le C-index de Harrell est défini par

$$\mathbb{P}(M_i > M_j \mid T_i < T_j)$$

où i et j sont deux individus indépendants. Autrement dit, il correspond à la probabilité que la mesure du marqueur de l’individu i soit strictement supérieure à celle de l’individu j sachant que le temps à l’évènement de i est strictement inférieur à celui de j .

En terme d’interprétation, plus le C-index est proche de 1 (ou 0), plus le marqueur associé est un marqueur pronostique de l’évènement étudié. En revanche, un C-index proche de 0.5 indiquera un mauvais pouvoir pronostique du marqueur. Contrairement à l’AUC dépendant du temps, le C-index n’est pas adapté à l’évaluation des prédictions à un horizon t fixé (BLANCHE, KATTAN et GERDS 2019). C’est un outil intéressant pour mesurer les capacités de discrimination globale d’un marqueur, mais l’AUC dépendant du temps doit lui être préféré pour en mesurer l’efficacité pronostique.

2.6.2 Évaluation des capacités de calibration d’un modèle : le Brier Score

Le Brier Score (BRIER 1950) est un outil permettant de mesurer simultanément la discrimination et la calibration. Il se présente sous la forme d’une erreur quadratique

moyenne entre ce qui a été observé et ce qui a été prédit par un modèle afin de valider les capacités prédictives de ce dernier.

Soient D_i une variable binaire observée indiquant si le patient i est malade ou non (1 si malade, 0 sinon), E_i la prédiction de risque du modèle pour l'individu i , $i \in \{1, \dots, N\}$, avec N le nombre total d'individus. Le Brier Score (BS) est alors défini par

$$BS = \frac{1}{N} \sum_{i=1}^N (E_i - D_i)^2.$$

Le Brier Score est compris entre 0 et 1 : plus il est proche de 0, meilleures sont les capacités prédictives du modèle. C'est un score populaire qui reste critiqué pour sa difficulté d'interprétation. En effet, en élevant au carré les erreurs comprises entre 0 et 1, les erreurs importantes auront moins de poids pour le Brier Score comparées numériquement aux erreurs plus faibles. Il doit être interprété avec précaution en essayant de comprendre si l'erreur est causée par un faible nombre de prédictions E_i avec une erreur élevée ou par un grand nombre de prédictions E_i avec une erreur plus faible.

Afin de différencier l'impact de la prévalence par rapport à celui des calibration et discrimination, une normalisation du Brier Score appelée *Index of Prediction Accuracy* (IPA) a été proposée (GRAF et al. 1999 ; KATTAN et GERDS 2018). En posant BS_0 le Brier score basé sur le modèle nul n'incluant aucune covariable, l'IPA est défini par

$$IPA = 1 - \frac{BS}{BS_0}.$$

Ces dernières années, des versions dynamiques de certains de ces critères permettant de prendre en compte le suivi des marqueurs ont été développées. Elles sont présentées dans la Section 2.6.3 suivante et permettent de déterminer si un marqueur est pertinent à suivre dans le temps afin de prédire les événements.

2.6.3 Évaluation de prédictions dynamiques

La médecine personnalisée est aujourd'hui un point fort de l'approche médicale qui accorde une place centrale à l'étude du pronostic individuel du patient. Le développement de la modélisation dite dynamique contribue à cet objectif. Elle permet de réaliser des prédictions dynamiques du risque individuel, elles-même mises à jour au fil du temps en fonction de l'évolution du profil clinique du patient. En particulier, les modèles conjoints

définis lors des Sections 2.5.1 et 2.5.2 ont l'avantage d'établir leurs prédictions en prenant en compte l'ensemble de l'historique médical du patient. Dans cette section, nous allons définir ces prédictions dynamiques et présenter deux critères de performance dynamiques permettant de les évaluer.

2.6.3.1 Prédictions dynamiques

Les prédictions dynamiques sont utilisées pour connaître la probabilité de risque pour un individu de subir un évènement sur une fenêtre de temps $]s; s + h]$, $s \in \mathbb{R}^+$, $h \in \mathbb{R}^{+*}$, sachant qu'il est toujours à risque au temps s et sachant tout l'historique des covariables mesurées avant s . C'est une modélisation landmark comme expliquée dans la Figure 2.5 de la Section 2.3.3. Elles peuvent être calculées à partir des modèles conjoints à effets aléatoires partagés (RIZOPOULOS 2011) ou à classes latentes (PROUST-LIMA et al. 2014).

Prédictions dynamiques à l'aide de modèles conjoints à effets aléatoires partagés En reprenant les notations de la Section 2.5.1, pour tout individu $i \in \{1, \dots, N\}$, la probabilité d'observer un évènement sur la fenêtre de temps $]s; s + h]$ est

$$\begin{aligned} & \mathbb{P}_{\hat{\theta}_{\text{JM}}} \left(T_i \leq s + h \mid T_i \geq s, \bar{Y}_i(s), \bar{X}_i(s) \right) \\ &= \int \mathbb{P}_{\hat{\theta}_{\text{JM}}} \left(T_i \leq s + h \mid T_i \geq s, b_i = u, \bar{X}_i(s) \right) f_{b_i} \left(u \mid T_i \geq s, \bar{Y}_i(s), \bar{X}_i(s) \right) du \\ &= \frac{\int f_{\bar{Y}_i(s)} \left(\bar{Y}_i(s) \mid b_i = u, \bar{X}_i(s); \hat{\theta}_{\text{JM}} \right) \left(S_{iu}(s; \hat{\theta}_{\text{JM}}) - S_{iu}(s + h; \hat{\theta}_{\text{JM}}) \right) f_{b_i} (u) du}{\int f_{\bar{Y}_i(s)} \left(\bar{Y}_i(s) \mid b_i = u, \bar{X}_i(s); \hat{\theta}_{\text{JM}} \right) S_{iu}(s; \hat{\theta}_{\text{JM}}) f_{b_i} (u) du} \end{aligned}$$

avec

- $\mathbb{P}_{\hat{\theta}_{\text{JM}}}$ la distribution de probabilité paramétrée par $\hat{\theta}_{\text{JM}}$, où $\hat{\theta}_{\text{JM}}$ est l'estimation de θ_{JM} ,
- $\bar{Y}_i(s) = \{Y_{ij}, j = 1, \dots, n_i, \text{ avec } t_{ij} < s\}$ les mesures de la variable aléatoire réponse jusqu'au temps s pour l'individu i ,
- $\bar{X}_i(s) = \{X_i^{\text{mxt}}, X_i^{\text{surv}}, j = 1, \dots, n_i, \text{ avec } t_{ij} < s\}$ les différentes covariables mesurées jusqu'au temps s ,
- T_i le temps à l'évènement, et $S_{iu}(s; \hat{\theta}_{\text{JM}}) = S_i \left(s \mid b_i = u, \bar{X}_i(s); \hat{\theta}_{\text{JM}} \right)$.

De la même manière, les prédictions définies dans le cadre de modèles conjoints à classes latentes sont présentées dans le paragraphe suivant.

Prédictions dynamiques à l'aide de modèles conjoints à classes latentes Les prédictions dynamiques issues des modèles conjoints à classes latentes sont similaires à celles obtenues avec les modèles à effets aléatoires partagés : l'intégrale sur les effets aléatoires est remplacée par une somme sur les classes latentes. En reprenant les notations des modèles à classes latentes définies en Section 2.5.2, pour tout individu $i \in \{1, \dots, N\}$, les prédictions dynamiques sont définies par

$$\begin{aligned} & \mathbb{P}_{\hat{\theta}_{\text{lcmmm}}} \left(T_i \leq s + h \mid T_i \geq s, \bar{Y}_i(s), \bar{X}_i(s) \right) \\ &= \sum_{g=1}^G \mathbb{P}_{\hat{\theta}_{\text{lcmmm}}} \left(T_i \leq s + h \mid T_i \geq s, c_i = g, \bar{X}_i(s) \right) \mathbb{P}_{\hat{\theta}_{\text{lcmmm}}} \left(c_i = g \mid T_i \geq s, \bar{Y}_i(s), \bar{X}_i(s) \right) \\ &= \frac{\sum_{g=1}^G \pi_{ig} f_{\bar{Y}_i(s)} \left(\bar{Y}_i(s) \mid c_i = g, \bar{X}_i(s); \hat{\theta}_{\text{lcmmm}} \right) \left(S_{ig}(s; \hat{\theta}_{\text{lcmmm}}) - S_{ig}(s + h; \hat{\theta}_{\text{lcmmm}}) \right)}{\sum_{g=1}^G \pi_{ig} f_{\bar{Y}_i(s)} \left(\bar{Y}_i(s) \mid c_i = g, \bar{X}_i(s); \hat{\theta}_{\text{lcmmm}} \right) S_{ig}(s; \hat{\theta}_{\text{lcmmm}})} \end{aligned}$$

avec

- $\mathbb{P}_{\hat{\theta}_{\text{lcmmm}}}$ la distribution de probabilité paramétrée par $\hat{\theta}_{\text{lcmmm}}$, où $\hat{\theta}_{\text{lcmmm}}$ est l'estimation de θ_{lcmmm} ,
- $\bar{Y}_i(s) = \{Y_{ij}, j = 1, \dots, n_i, \text{ avec } t_{ij} < s\}$ les mesures de la variable aléatoire réponse jusqu'au temps s pour l'individu i ,
- $\bar{X}_i(s) = \{X_i^{\text{mlt}}, X_i^{\text{mxt}}, X_i^{\text{surv}}, j = 1, \dots, n_i, \text{ avec } t_{ij} < s\}$ les différentes covariables mesurées jusqu'au temps s ,
- et $S_{iu}(s; \hat{\theta}_{\text{lcmmm}}) = S_i(s \mid c_i = g, \bar{X}_i(s); \hat{\theta}_{\text{lcmmm}})$.

Exemple 11. *Cet exemple présente l'évolution de prédictions dynamiques calculées à l'aide d'un modèle conjoint à classes latentes durant le suivi d'un individu. Pour cela, considérons un marqueur longitudinal mesuré à plusieurs temps durant le suivi d'un individu. Les prédictions dynamiques sont calculées aux temps landmark $s = 2$ et $s = 4$ ans avec les horizons $h = \{1, 2, 3, 4\}$ ans. La Figure 2.10 présente les différentes prédictions dynamiques obtenues.*

Sur la Figure 2.10, l'axe des abscisses représente le temps, l'axe des ordonnées à gauche indique la valeur du marqueur longitudinal, et l'axe des ordonnées à droite la probabilité pour l'individu de subir l'évènement. Chaque point correspond à une nouvelle mesure du

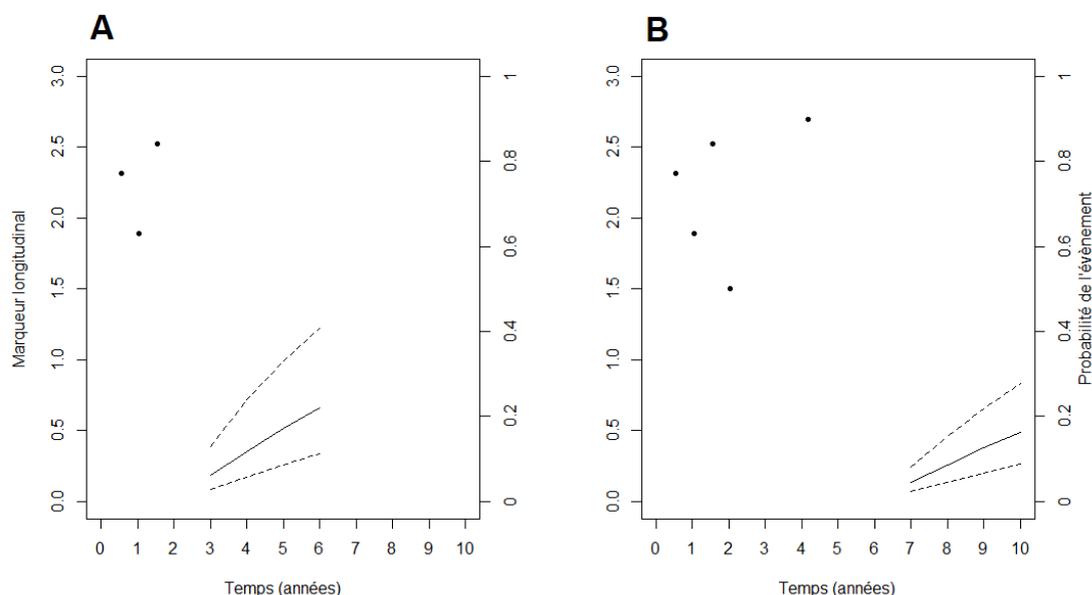


FIGURE 2.10 – Prédications dynamiques calculées aux temps landmark $s = 2$ (A) et $s = 6$ ans (B) pour des horizons de $h = \{1, 2, 3, 4\}$ ans

marqueur. Les prédictions dynamiques sont représentées par le trait plein avec l'intervalle de confiance à 95% associé en pointillés.

La probabilité pour un individu de subir un évènement sur la fenêtre de temps $]2; 3]$, sachant qu'il est toujours à risque à $s = 2$ ans et sachant tout l'historique des covariables mesurées avant $s = 2$ ans, est de 6%. Elle égale à 12% pour la fenêtre d'observation $]2; 4]$, 17% pour $]2; 5]$ et 22% pour $]2; 6]$. Ces prédictions sont basées sur les trois premières mesures ayant lieu durant les deux premières années de suivi. En récoltant plus de suivi pour l'individu et en se plaçant au temps landmark $s = 6$ ans, la probabilité de subir l'évènement est de 4% entre 6 et 7 ans, 9% entre 6 et 8 ans, 13% entre 6 et 9 ans, et 16% entre 6 et 10 ans. Ici, les prédictions dynamiques tiennent toujours compte des trois premières mesures et sont complétées par deux nouvelles à 2 et 4 ans.

Pour évaluer les capacités de ces prédictions dynamiques, nous utiliserons deux critères : l'AUC et le Brier Score. Comme déjà dit précédemment, ces critères sont utilisés en statistiques pour mesurer la discrimination et la calibration. Appropriés pour l'évaluation pronostique (GERDS, CAI et SCHUMACHER 2008), les versions dynamiques de ces deux critères sont présentées dans la Section 2.6.3.2 suivante.

2.6.3.2 Critères de performance dynamiques

En fixant le temps landmark $s \in \mathbb{R}^+$ et l'horizon $h \in \mathbb{R}^{+*}$, la capacité de ces prédictions à prévoir correctement les événements survenant entre s et $s+h$ peut être évaluée à l'aide de l'AUC et du Brier Score dont des versions dynamiques ont été développées (BLANCHE et al. 2015). En reprenant les notations de la Section 2.2.2.2, soient $\tilde{T} = \min(T, C_d)$ le temps d'observation et $\delta = \mathbb{1}_{\{T \leq C_d\}}$ l'indicateur d'évènement, avec T la variable de temps à l'évènement et C_d le temps de censure à droite. Pour tout $i \in \{1, \dots, N\}$, la prédiction dynamique du risque de l'évènement entre s et $s+h$ pour l'individu i est donnée par

$$\pi_i^{\text{dynpred}}(s, h) = \mathbb{P}(T_i \leq s+h \mid T_i \geq s, \bar{Y}_i(s), \bar{X}_i(s)),$$

avec $\bar{Y}_i(s)$ (respectivement $\bar{X}_i(s)$) l'ensemble des mesures répétées de la variable Y (respectivement des covariables X) jusqu'au temps s pour l'individu i . Elle est obtenue à l'aide d'un modèle conjoint à effets aléatoires partagés ou à classes latentes.

AUC dynamique L'AUC dynamique est utilisé afin de mesurer la discrimination des prédictions dynamiques. Soit $D_i(s, h) = \mathbb{1}_{(s < T_i \leq s+h)}$ égal à 1 si l'individu i subit l'évènement entre $]s, s+h]$, 0 sinon. Pour tout $i, j \in \{1, \dots, N\}$, l'AUC dynamique de BLANCHE et al. 2015 au temps landmark s et à l'horizon h est défini par

$$AUC(s, h) = \mathbb{P}(\pi_i^{\text{dynpred}}(s, h) > \pi_j^{\text{dynpred}}(s, h) \mid D_i(s, h) = 1, D_j(s, h) = 0, T_i > s, T_j > s).$$

Il est égal à la probabilité que l'individu i possède une probabilité de risque plus grande que l'individu j , sachant que i a subi l'évènement entre s et $s+h$ mais pas j , et qu'ils sont encore tous les deux à risque au temps s .

Afin de prendre en compte la censure lors de l'estimation, UNO et al. 2007 et HUNG et CHIANG 2010 ont développé des estimateurs *Inverse Probability of Censoring Weighting* (IPCW). Ce type d'estimateur consiste à pondérer les individus par la probabilité de ne pas être censuré. En supposant le temps de censure C_d indépendant de $(T, \pi^{\text{dynpred}}(\cdot, \cdot))$, les poids sont définis de la façon suivante

$$\widehat{W}_i(s, h) = \frac{\mathbb{1}_{(\tilde{T}_i > s+h)}}{\widehat{G}(s+h \mid s)} + \frac{\mathbb{1}_{(s < \tilde{T}_i \leq s+h)} \Delta_i}{\widehat{G}(\tilde{T}_i \mid s)},$$

où $\widehat{G}(u)$ est l'estimateur de Kaplan-Meier de $\mathbb{P}(C_d > u)$, et $\widehat{G}(u | s) = \widehat{G}(u)/\widehat{G}(s)$ la probabilité de ne pas être censuré à u sachant qu'on ne l'est pas à s . Soit $\widetilde{D}_i(s, h) = \mathbb{1}_{(s < \widetilde{T}_i \leq s+h)}$. Ainsi, BLANCHE et al. 2015 proposent d'estimer $AUC(s, h)$ par

$$\widehat{AUC}(s, h) = \frac{\sum_{i=1}^n \sum_{j=1}^n \mathbb{1}_{(\pi_i^{\text{dynpred}}(s, h) > \pi_j^{\text{dynpred}}(s, h))} \widetilde{D}_i(s, h) (1 - \widetilde{D}_j(s, h)) \widehat{W}_i(s, h) \widehat{W}_j(s, h)}{\sum_{i=1}^n \sum_{j=1}^n \widetilde{D}_i(s, h) (1 - \widetilde{D}_j(s, h)) \widehat{W}_i(s, h) \widehat{W}_j(s, h)}.$$

Brier Score dynamique Afin de mesurer la discrimination et la calibration de prédictions dynamiques, le Brier Score dynamique a été développé (BLANCHE et al. 2015) :

$$BS(s, h) = \mathbb{E} \left[\left(D(s, h) - \pi^{\text{dynpred}}(s, h) \right)^2 \mid T > s \right].$$

Cette définition s'appuie sur celle de SCHOOP et al. 2011. Il correspond à une erreur quadratique entre ce qui a été observé et ce qui a été prédit, et est estimé par

$$\widehat{BS}(s, h) = \frac{1}{N \widehat{S}_{\widetilde{T}}(s)} \sum_{i=1}^N \widehat{W}_i(s, h) \left(\widetilde{D}_i(s, h) - \pi_i^{\text{dynpred}}(s, h) \right)^2,$$

où $\widehat{S}_{\widetilde{T}}(s)$ est la probabilité estimée d'observer un individu à risque au temps s .

Un test de comparaison entre deux critères de performance à un temps landmark donnée ou sur un ensemble de temps landmark a été développé par BLANCHE et al. 2015. Cependant, seuls deux critères peuvent être comparés l'un à l'autre. L'objectif du Chapitre 5 est de développer une procédure afin de comparer plus de deux critères de performance entre eux. Pour cela, plusieurs tests sont effectués tout en corrigeant la multiplicité. En effet, plus le nombre de tests est important, plus le risque de faire une fausse découverte augmente. Une correction des p-valeurs est alors indispensable grâce à des procédures de gestion de la multiplicité des tests dont une introduction est présentée dans la Section 2.7 suivante.

2.7 Multiplicité des tests

Quand une seule hypothèse nulle est impliquée, les tests statistiques sont souvent choisis pour contrôler le nombre de rejets à tort de \mathcal{H}_0 à un certain seuil de confiance. Cepen-

dant, ce nombre d'erreurs augmente lorsque plusieurs hypothèses nulles sont testées : les p-valeurs doivent être corrigées pour ne pas faire de fausses découvertes et ainsi conclure à une différence lorsqu'il n'y en a pas. Dans cette section, il est question de présenter deux types d'erreurs généralisés au contexte des tests multiples que l'on peut rencontrer lors de tests statistiques, ainsi que des méthodes de correction de la multiplicité.

2.7.1 Taux d'erreur

2.7.1.1 Erreur de type I

Un premier type d'erreur lié à l'échantillonnage aléatoire est l'erreur de type I notée α . Elle correspond à la probabilité de rejeter à tort l'hypothèse nulle :

$$\alpha = \mathbb{P}(\text{« rejeter } \mathcal{H}_0 \text{ »} \mid \text{« } \mathcal{H}_0 \text{ vraie »}).$$

Une approche classique est de choisir un test qui maintient le taux d'erreur de type I à un niveau de signification α préalablement choisi (généralement 0.05). Différentes extensions du taux d'erreur de type I aux problèmes de tests multiples sont possibles.

Soit $\varrho^{\text{rejet.faux}} \in \mathbb{N}$ le nombre d'hypothèses nulles rejetées à tort. Le *Family-Wise Error Rate* (FWER) est le taux d'erreur le plus couramment utilisé lors de tests multiples, en particulier lorsque le nombre de tests est modéré ou lors d'analyses confirmatoires (BRETZ, HOTHORN et WESTFALL 2016, p. 13). Il est défini par

$$FWER = \mathbb{P}(\varrho^{\text{rejet.faux}} > 0)$$

et correspond à la probabilité de rejeter au moins une fois à tort l'hypothèse nulle. Le FWER est dit contrôlé à un niveau α s'il est inférieur ou égal à α . Une autre façon de contrôler l'erreur de type I a été développée par BENJAMINI et HOCHBERG 1995. Elle est appelée *False Discovery Rate* (FDR) et s'écrit

$$FDR = \mathbb{E} \left[\frac{\varrho^{\text{rejet.faux}}}{\varrho^{\text{rejet}}} \mid \varrho^{\text{rejet}} > 0 \right] \times \mathbb{P}(\varrho^{\text{rejet}} > 0),$$

avec $\varrho^{\text{rejet}} \in \mathbb{N}$ le nombre total d'hypothèses nulles rejetées. Principalement utilisé dans un cadre exploratoire, le FDR correspond à l'espérance de la proportion d'erreurs de type I parmi les hypothèses nulles rejetées.

2.7.1.2 Erreur de type II

Le deuxième type d’erreur rencontré est l’erreur de type II. Elle est notée β et est définie comme la probabilité de ne pas rejeter \mathcal{H}_0 sachant que \mathcal{H}_0 est fausse. Le complémentaire de β est appelée la puissance :

$$1 - \beta = \mathbb{P}(\text{« conclure } \mathcal{H}_1 \text{ »} \mid \text{« } \mathcal{H}_1 \text{ vraie »}).$$

En d’autres termes, la puissance correspond à la probabilité de conclure à une différence sachant qu’il y en a une. Le but pour chaque test statistique est de maximiser la puissance et de minimiser le taux d’erreur de type II pour un critère d’erreur de type I donné. Comme pour l’extension du taux d’erreur de type I au FWER ou FDR, la puissance peut être généralisée de diverses manières lorsqu’on passe dans le cadre des tests multiples.

Soient $v \in \mathbb{N}$ le nombre de fausses hypothèses nulles, et $\varrho^{\text{rejet.vrai}} \in \mathbb{N}$ le nombre de fausses hypothèses nulles rejetées à raison. La première puissance présentée ici est la puissance disjonctive (BRETZ, HOTHORN et WESTFALL 2016, p. 16). Également appelée puissance minimale (MAURER et MELLEIN 1988), elle s’écrit

$$\mathbb{P}(\varrho^{\text{rejet.vrai}} \geq 1)$$

et correspond à la probabilité de rejeter au moins une fausse hypothèse nulle. Un second type de puissance utilisé est la puissance conjonctive (BRETZ, HOTHORN et WESTFALL 2016, p. 16) :

$$\mathbb{P}(\varrho^{\text{rejet.vrai}} = v).$$

Aussi appelée puissance totale ou complète (WESTFALL et al. 1999), elle est définie comme la probabilité de rejeter toutes les hypothèses nulles. Enfin, la *r-power* (DUNNETT et TAMHANE 1992) est la dernière puissance présentée dans cette section. Elle représente la probabilité de rejeter au moins $r \in \mathbb{R}$ fausses hypothèses nulles à raison et s’écrit

$$\mathbb{P}(\varrho^{\text{rejet.vrai}} \geq r).$$

Dans le contexte des tests multiples, la problématique est donc de minimiser les erreurs de type II tout en contrôlant le taux d’hypothèses nulles rejetées à tort. Cependant, contrôler les erreurs de type I à un niveau α pour chacune des hypothèses individuelles

ne suffit pas pour contrôler le FWER ou le FDR à un niveau α . Dans le cadre d'analyses exploratoires, c'est le FDR qui est choisi pour contrôler l'erreur de type I. Moins contraignant que le FWER, il tolère une certaine proportion d'hypothèses nulles rejetées à tort. Utilisé lors d'analyses confirmatoires où une plus grande précision est exigée, le FWER réduit le nombre d'erreurs aux dépens de la puissance. Afin de contrôler les types d'erreurs de type I présentés, des procédures de correction des p-valeurs ont été mises en place afin de gérer la multiplicité des tests.

2.7.2 Correction de la multiplicité

Soient $K \geq 2$ hypothèses nulles individuelles $\mathcal{H}_0^k, k \in \{1, \dots, K\}$, testées par une procédure de niveau de α , et T_{stat}^k les K statistiques de test associées. Soient $p^k = \mathbb{P}(T_{\text{stat}}^k > t_{\text{stat}}^k)$ les p-valeurs associées à T_{stat}^k où t_{stat}^k est la réalisation de T_{stat}^k . Plusieurs procédures existent afin de contrôler les erreurs de type I. Dans cette section, celles de Bonferroni, Bonferroni-Holm et Shaffer figurant parmi les plus populaires sont présentées.

Procédure de Bonferroni La procédure la plus naïve et la plus facile à utiliser est la correction de Bonferroni (BONFERRONI 1935). Pour tout $k \in \{1, \dots, K\}$, elle consiste à rejeter \mathcal{H}_0^k si $p^k \leq \alpha/K$. Cependant, elle est connue pour être très conservatrice lorsque le nombre d'hypothèses à tester augmente ou que les hypothèses à tester sont corrélées.

Procédure de Bonferroni-Holm La procédure de Bonferroni a été améliorée par HOLM 1979 en appliquant le principe du *closed testing* (MARCUS, ERIC et GABRIEL 1976). Elle est appelée procédure de Bonferroni-Holm et corrige les p-valeurs p^k de manière séquentielle. Soient $\{p^{(k)}\}$ les K p-valeurs ordonnées de p^k telles que $p^{(1)} \leq \dots \leq p^{(K)}$, et $\mathcal{H}_0^{(k)}$ les hypothèses liées au test (k) . L'hypothèse nulle $\mathcal{H}_0^{(1)}$ est rejetée si $p^{(1)} \leq \alpha/K$. Pour tout $k \in \{2, \dots, K\}$, $\mathcal{H}_0^{(k)}$ est rejetée si $\mathcal{H}_0^{(k-1)}$ l'est aussi et si $p^{(k)} \leq \alpha/(K - k + 1)$. Ainsi, ne pas rejeter $\mathcal{H}_0^{(k)}, k \in \{1, \dots, K\}$, implique de ne pas rejeter $\mathcal{H}_0^{(l)}, 1 \leq k < l \leq K$. La procédure de Bonferroni-Holm possède une meilleure puissance que celle de Bonferroni.

Procédure de Shaffer Enfin, SHAFFER 1986 a amélioré la procédure de Bonferroni-Holm : seul un sous-ensemble d'hypothèses est testé en utilisant des contraintes logiques. Au lieu de rejeter $\mathcal{H}_0^{(k)}, k \in \{1, \dots, K\}$ si $p^{(k)} \leq \alpha/(K - k + 1)$ (procédure de Bonferroni-Holm), la procédure de Shaffer rejette $\mathcal{H}_0^{(k)}$ si $p^{(k)} \leq \alpha/m_{\text{shaffer}}$, avec m_{shaffer} le nombre maximum d'hypothèses vraies possibles, sachant qu'au moins $(k - 1)$ hypothèses sont

fausses. Lorsqu'il existe des relations logiques entre les hypothèses, le nombre d'hypothèses vraies ne peut pas prendre certaines valeurs entre 0 et K : le fait que $(k - 1)$ hypothèses soient fausses implique que d'autres le soient pour certains k . Pour ces valeurs de k , m_{shaffer} sera strictement inférieur à $(K - k + 1)$. Comme m_{shaffer} ne peut être supérieur à $(K - k + 1)$, la procédure de Shaffer sera au moins aussi puissante que la procédure Bonferroni-Holm. Dans la plupart des applications avec des combinaisons restreintes, elle sera uniformément plus puissante.

STRATIFICATION DU RISQUE DE COMPLICATIONS HÉPATIQUES PAR DES TESTS NON-INVASIFS CHEZ LES PATIENTS ATTEINTS DE NAFLD

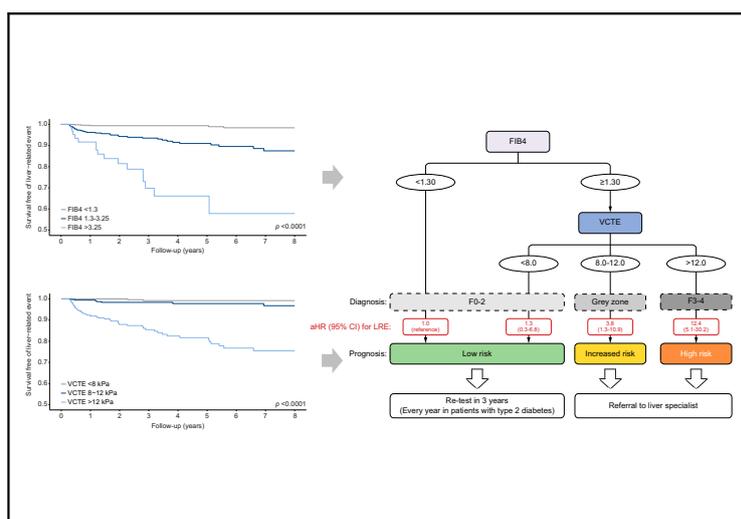
Les études précédentes sur la signification pronostique des tests non-invasifs de fibrose hépatique dans la *Non-Alcoholic Fatty Liver Disease* (NAFLD) ne sont pas directement comparables à la biopsie du foie. Dans ce chapitre, nous avons cherché à évaluer la précision pronostique d'un test sanguin (FIB-4) et du Fibroscan (aussi appelé VCTE pour *Vibration-Controlled Transient Elastography*) pour la prédiction des événements hépatiques dans la NAFLD par rapport à la biopsie du foie. Ces résultats ont été publiés dans *Journal of Hepatology* en janvier 2022.

3.1 Article scientifique

Résumé L'importance de la fibrose dans le foie est étroitement liée au risque de complications hépatiques dans la NAFLD. La biopsie hépatique reste actuellement la norme de référence pour l'évaluation de la fibrose, mais son application est limitée par son caractère invasif. Nous avons donc évalué la capacité des tests non-invasifs de fibrose hépatique à prédire les complications liées au foie dans la NAFLD. Nos résultats montrent que le test sanguin FIB-4 et la mesure de l'élasticité du foie par le VCTE stratifient le risque de complications hépatiques dans la NAFLD. De plus, le VCTE possède des capacités pronostiques similaires à celles de la biopsie hépatique. Ces résultats confortent l'utilisation de tests non-invasifs de fibrose hépatique au lieu de la biopsie pour la gestion des patients atteints de NAFLD.

Non-invasive tests accurately stratify patients with NAFLD based on their risk of liver-related events

Graphical abstract



Highlights

- At their published cut-offs, FIB4 and VCTE stratify patients with NAFLD into subgroups with significantly different prognoses.
- Compared to liver biopsy in NAFLD, VCTE has similar accuracy for the prediction of liver-related events.
- With regards to clinical events, the FIB4-VCTE stepwise algorithm accurately discriminates at-risk patients with NAFLD.
- This algorithm should be used in the referral pathway to a liver specialist.

Authors

Jerome Boursier, Hannes Hagström, Mattias Ekstedt, ..., Jeremie Riou, Aldo Trylesinski, Manuel Romero-Gomez

Correspondence

JeBoursier@chu-angers.fr (J. Boursier).

Lay summary

The amount of fibrosis in the liver is closely associated with the risk of liver-related complications in non-alcoholic fatty liver disease (NAFLD). Liver biopsy currently remains the reference standard for the evaluation of fibrosis, but its application is limited by its invasiveness. Therefore, we evaluated the ability of non-invasive liver fibrosis tests to predict liver-related complications in NAFLD. Our results show that the blood test FIB4 and transient elastography stratify the risk of liver-related complications in NAFLD, and that transient elastography has similar prognostic accuracy as liver biopsy. These results support the use of non-invasive liver fibrosis tests instead of liver biopsy for the management of patients with NAFLD.

<https://doi.org/10.1016/j.jhep.2021.12.031>

© 2022 European Association for the Study of the Liver. Published by Elsevier B.V. All rights reserved. J. Hepatol. 2022, ■, 1–8

Non-invasive tests accurately stratify patients with NAFLD based on their risk of liver-related events

Jerome Boursier^{1,2,*}, Hannes Hagström^{3,4,5}, Mattias Ekstedt⁶, Clemence Moreau², Martin Bonacci^{7,8}, Sandrine Cure^{7,8}, Javier Ampuero⁹, Patrik Nasr⁶, Lilian Tallab⁴, Clémence M. Canivet^{1,2}, Stergios Kechagias⁶, Yolanda Sánchez⁹, Eloise Dincuff¹, Ana Lucena⁹, Marine Roux², Jeremie Riou², Aldo Trylesinski^{7,8}, Manuel Romero-Gomez⁹

¹Service d'Hépatogastroentérologie et Oncologie Digestive, Centre Hospitalier Universitaire d'Angers, Angers, France; ²Laboratoire HIFIH UPRES EA3859, SFR ICAT 4208, Université d'Angers, Angers, France; ³Unit of Hepatology, Department of Upper GI, Karolinska University Hospital, Stockholm, Sweden; ⁴Department of Medicine, Huddinge, Karolinska Institutet, Stockholm, Sweden; ⁵Unit of Clinical Epidemiology, Department of Medicine, Solna, Karolinska Institutet, Stockholm, Sweden; ⁶Department of Health, Medicine and Caring Sciences, Division of Diagnostics and Specialist Medicine, Linköping University, Linköping, Sweden; ⁷Intercept Pharmaceuticals, London, UK; ⁸Intercept Pharmaceuticals, Inc., New York, NY, USA; ⁹Digestive Diseases Department, Virgen del Rocío University Hospital, Institute of Biomedicine of Seville, University of Seville, Seville, Spain

Background & Aims: Previous studies on the prognostic significance of non-invasive liver fibrosis tests in non-alcoholic fatty liver disease (NAFLD) lack direct comparison to liver biopsy. We aimed to evaluate the prognostic accuracy of fibrosis-4 (FIB4) and vibration-controlled transient elastography (VCTE), compared to liver biopsy, for the prediction of liver-related events (LREs) in NAFLD.

Methods: A total of 1,057 patients with NAFLD and baseline FIB4 and VCTE were included in a multicenter cohort. Of these patients, 594 also had a baseline liver biopsy. The main study outcome during follow-up was occurrence of LREs, a composite endpoint combining cirrhosis complications and/or hepatocellular carcinoma. Discriminative ability was evaluated using Harrell's C-index.

Results: FIB4 and VCTE showed good accuracy for the prediction of LREs, with Harrell's C-indexes >0.80 (0.817 [0.768-0.866] vs. 0.878 [0.835-0.921], respectively, $p = 0.059$). In the biopsy subgroup, Harrell's C-indexes of histological fibrosis staging and VCTE were not significantly different (0.932 [0.910-0.955] vs. 0.881 [0.832-0.931], respectively, $p = 0.164$), while both significantly outperformed FIB4 for the prediction of LREs. FIB4 and VCTE were independent predictors of LREs in the whole study cohort. The stepwise FIB4-VCTE algorithm accurately stratified the risk of LREs: compared to patients with "FIB4 <1.30", those with "FIB4 \geq 1.30 then VCTE <8.0 kPa" had similar risk of LREs (adjusted hazard ratio [aHR] 1.3; 95% CI 0.3-6.8), whereas the risk of LREs significantly increased in patients with "FIB4 \geq 1.30 then VCTE 8.0-12.0 kPa" (aHR 3.8; 95% CI 1.3-10.9), and even more for those with "FIB4 \geq 1.30 then VCTE >12.0 kPa" (aHR 12.4; 95% CI 5.1-30.2).

Conclusion: VCTE and FIB4 accurately stratify patients with NAFLD based on their risk of LREs. These non-invasive tests are alternatives to liver biopsy for the identification of patients in need of specialized management.

Lay summary: The amount of fibrosis in the liver is closely associated with the risk of liver-related complications in non-alcoholic fatty liver disease (NAFLD). Liver biopsy currently remains the reference standard for the evaluation of fibrosis, but its application is limited by its invasiveness. Therefore, we evaluated the ability of non-invasive liver fibrosis tests to predict liver-related complications in NAFLD. Our results show that the blood test FIB4 and transient elastography stratify the risk of liver-related complications in NAFLD, and that transient elastography has similar prognostic accuracy as liver biopsy. These results support the use of non-invasive liver fibrosis tests instead of liver biopsy for the management of patients with NAFLD.

© 2022 European Association for the Study of the Liver. Published by Elsevier B.V. All rights reserved.

Introduction

Among all liver histology features observed in non-alcoholic fatty liver disease (NAFLD), liver fibrosis is the main determinant of patient prognosis.¹ Longitudinal studies have demonstrated that the risk of liver-related events (LREs) significantly increases at fibrosis stage 2 (F2, significant fibrosis) and becomes exponentially higher when transitioning to stage F3 (bridging fibrosis) and later F4 (cirrhosis).¹ Therefore, liver fibrosis should be evaluated in all patients with NAFLD to determine the stage of the disease and to guide patient management, including the allocation of treatments when they become available.²⁻⁴

Liver biopsy remains the reference for evaluating disease severity in NAFLD, but its invasive nature, several pragmatic issues (e.g., associated costs), limitations to scalability, and finally the patient's reluctance to undergo the procedure limit its widespread use for that purpose. Furthermore, liver biopsy is still hampered by sample bias and significant limitations in inter-observer reproducibility.^{5,6} Several non-invasive tests, mainly elastography devices and blood marker-based algorithms, are

Keywords: NAFLD; fibrosis; FIB4; Fibroscan; prognosis; liver-related event.

Received 14 March 2021; received in revised form 20 November 2021; accepted 23 December 2021; available online xxx

* Corresponding author. Address: 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 address: JeBoursier@chu-angers.fr (J. Boursier).

<https://doi.org/10.1016/j.jhep.2021.12.031>



ELSEVIER

now available to estimate liver fibrosis. These tests have demonstrated good accuracy for the diagnosis of advanced fibrosis in NAFLD.⁷⁻⁹ However, diagnostic studies use liver biopsy as gold standard, when in reality it is an imperfect reference leading to underestimation of the true performance of non-invasive tests.¹⁰ Moreover, reaching perfect diagnostic accuracy for the non-invasive staging of liver fibrosis is probably an unattainable goal as a liver architecture semi-quantitative classification (histology) cannot perfectly correlate with a quantitative measurement such as blood protein levels (blood tests) or a physical signal such as the speed of a shear wave in the liver (elastography).

An interesting concept has recently emerged though. Considering that non-invasive tests correlate to liver fibrosis, and fibrosis predicts prognosis, non-invasive tests may work as prognostic markers in patients with NAFLD.¹¹⁻¹⁴ This approach could change the paradigm in NAFLD management, moving from the diagnosis of liver lesions to the prediction of prognosis. Resulting from this concept, a strong argument for the use of non-invasive tests instead of liver biopsy when discussing NAFLD management in clinical practice would be to demonstrate that these different modalities provide similar information regarding the assessment of prognosis.

In the study presented here, we aimed to evaluate the prognostic significance of non-invasive liver fibrosis tests with comparison to histological fibrosis staging in a multicenter cohort of patients with NAFLD.

Patients and methods

Patients

Local cohorts of patients with NAFLD from 5 tertiary centers (Angers in France, Linköping and Stockholm in Sweden, Virgen del Rocío and Valme Hospitals in Seville, Spain) were pooled for the present study (details about the cohorts are presented in the [supplementary patients and methods](#)). NAFLD was defined as the presence of liver steatosis on imaging or histology, without any of the following conditions: concomitant steatosis-inducing drugs, excessive alcohol consumption (>210 g/week in men or >140 g/week in women), or other causes of chronic liver disease (chronic viral hepatitis, hemochromatosis, cholestatic chronic liver disease, auto-immune hepatitis, alpha-1 antitrypsin deficiency). Patients with baseline fibrosis-4 (FIB4) blood test and/or vibration-controlled transient elastography (VCTE) were included in the present work. Those with a history of bariatric surgery, a history of LREs before baseline, and/or follow-up <3 months were excluded. The study protocol conformed to the ethical guidelines of the current Declaration of Helsinki and received approval by the local Ethics Committees.

Non-invasive tests of liver fibrosis

Clinical data, biology, and liver stiffness results were collected at baseline and follow-up visits. Recorded clinical data and blood parameters were age, sex, antidiabetic treatment, antihypertensive treatment, lipid-lowering treatment, aspartate aminotransferase (AST, IU/L), alanine aminotransferase (ALT, IU/L), gamma-glutamyl transferase (IU/L) and platelets ($10^9/L$). Diabetes was defined as the use of antidiabetic medications. The blood fibrosis test FIB4 was calculated according to the published formula: $[\text{age} * \text{AST} [\text{IU/L}] / \{\text{platelets} (10^9/L) * \text{ALT} (\text{IU/L})^{1/2}\}]$.¹⁵ Published cut-offs of 1.30 and 3.25 were used to categorize patients into subgroups based on their FIB4 results.^{15,16} To avoid unreliable results

linked to extrahepatic affections, FIB4 results were considered invalid and excluded from the statistical analysis if any of the following conditions were met: AST ≥ 400 IU/L, ALT ≥ 400 IU/L, platelets $\leq 20 \cdot 10^9/L$ or $\geq 500 \cdot 10^9/L$.

Liver stiffness measurements were performed using VCTE technology (FibroScan device; Echosens, Paris, France). The examinations were performed by experienced observers according to the manufacturer's recommendations.¹⁷ Ten valid measurements were required and the VCTE results were expressed in kPa as the median of those valid measurements. Ninety-two percent of the VCTE examinations were reliable according to the criteria derived from diagnostic studies.⁸ Patients were categorized into 3 groups based on their VCTE results, according to cut-offs recently published in a large multicenter study¹⁸ and a meta-analysis¹⁹: <8.0 kPa (low risk of advanced fibrosis), 8.0–12.0 kPa (intermediate risk of advanced fibrosis), and >12.0 kPa (high risk of advanced fibrosis).

The blood test FIB4 and VCTE were combined in the FIB4-VCTE algorithm (Fig. S1). This combinatorial stepwise approach was initially developed for the non-invasive diagnosis of advanced liver fibrosis,^{8,20} and has been recently proposed as a pathway for patient referral to liver specialists.²¹

Liver biopsy

Data on liver biopsy were collected when available at baseline. Liver histological lesions were evaluated in each center by a senior expert specialized in hepatology as per the NASH CRN classification.²² Inter-observer agreement for liver fibrosis staging in NAFLD has been shown to be very good to excellent between expert pathologists from tertiary centers.²²⁻²⁴ Non-alcoholic steatohepatitis (NASH) was defined as the presence of all 3 of the following conditions: steatosis grade ≥ 1 , lobular inflammation grade ≥ 1 , and ballooning grade ≥ 1 . Fibrosis was staged as follows: F0 = no fibrosis, F1 = perisinusoidal or portal/periportal fibrosis, F2 = perisinusoidal and portal/periportal fibrosis, F3 = bridging fibrosis and F4 = cirrhosis. "Advanced fibrosis" was defined as fibrosis stage $F \geq 3$, and "no/mild fibrosis" as F0–2.

Follow-up and study outcomes

Clinical events were retrieved from patient files according to a predefined procedure to ensure that data collection was conducted in a similar manner across participating centers. Clinical events recorded during follow-up were cirrhosis complications (*i.e.*, clinical ascites, spontaneous bacterial peritonitis, hepatorenal syndrome, gastrointestinal varices, variceal bleeding, liver failure, and encephalopathy [absent / present regardless of the severity stage]), hepatocellular carcinoma, and death. The primary outcome of the study was LREs, a composite endpoint defined by the first occurrence of a cirrhosis complication or hepatocellular carcinoma. Patients were followed until the first LRE, death, or the last available follow-up visit. In cases of bariatric surgery, patients were censored as without LREs at the date of that surgery.

Statistics

Missing data were handled using multiple imputation by random forests with the R package *missForest*.²⁵ Continuous variables were expressed as medians with 1st and 3rd quartiles, and categorical data were expressed as total numbers and percentages. The ability of the fibrosis tests to discriminate between patients

with and without future LREs was evaluated using Harrell's C-index.²⁶ Harrell's C-indexes were computed and compared using the R package *compareC*, according to methodology previously published.²⁷ Multiple testing was corrected using the Bonferroni method. Survival curves were determined using the Kaplan-Meier method and compared with the log-rank test. To evaluate the prognostic significance of the non-invasive tests used with their published cut-offs, we performed multivariable Cox models including FIB4 or VCTE with adjustment for age, sex, antidiabetic treatment, antihypertensive treatment, and lipid-lowering treatment. For these analyses, the non-invasive test was introduced as a categorical variable, *i.e.*, 3 groups defined by the published cut-offs (FIB4: <1.30, 1.30-3.25, >3.25; VCTE: <8.0 kPa, 8.0-12.0 kPa, >12.0 kPa), with calculation of the corresponding adjusted hazard ratios (aHRs) for LREs.

A competing risk regression considering death as the competing event for LREs was performed using the Fine-Gray method.²⁸ Survival free of LREs with the Kaplan-Meier method and with the cumulative incidence function considering death as competing risk (R package *cmprsk*) were computed on the same graph. Survival curves with both methods were perfectly superimposed (Fig. S2), which showed that death as a competing risk did not influence the results of our study. Therefore, results from the most parsimonious analyses (standard Cox model) are presented in the main manuscript, and those from the competing risk analyses in the [supplementary information](#). All statistical analyses were performed using R version 3.6.2.

Results

Patients

In all, 1,376 patients with NAFLD were screened for the study. After removing those with exclusion criteria (Fig. S3), 1,057 patients with baseline FIB4 and VCTE were included, of whom 594 had a liver biopsy available within 1 year from baseline (biopsy group). Baseline characteristics of the 1,057 patients included are presented in Table 1. Median age was 55 years, 62% of the patients were male, 37% had diabetes, 44% were under

antihypertensive treatment, and 27% were under lipid-lowering treatment. LREs occurred in 62 patients during the 3.1 years median follow-up (Table 2). The first occurrence was hepatocellular carcinoma in 14 patients and cirrhosis complications in the other 48 (ascites: n = 26, gastrointestinal varices: n = 18, liver failure/encephalopathy: n = 4). Among the 32 deaths observed during the patient follow-up, 17 occurred after LREs.

Comparison of prognostic accuracy between fibrosis tests

FIB4 and VCTE (both as continuous variables) showed good accuracy for the prediction of LREs in the full study cohort, with Harrell's C-indexes above 0.80 (Table 3). VCTE had higher Harrell's C-index than FIB4, but the difference did not reach statistical significance (0.878 [0.835-0.921] vs. 0.817 [0.768-0.866], $p = 0.059$). In the biopsy subgroup, Harrell's C-indexes of liver biopsy and VCTE were not significantly different; however, both significantly outperformed FIB4 for the prediction of LREs (Table 3). With regards to the prediction of all-cause mortality in the whole study cohort, FIB4 showed better discrimination than VCTE (0.895 [0.858-0.932] vs. 0.717 [0.590-0.844], $p = 0.003$).

Prognostic validation of the published diagnostic cut-offs for FIB4 and VCTE

Patients were categorized into 3 subgroups based on their baseline FIB4 and VCTE results (according to published cut-offs). LRE-free survival according to baseline non-invasive tests in the whole study cohort is depicted in Fig. 1. Univariate analysis showed that subgroups defined by the non-invasive tests had significantly different prognoses (Table 4). Patients with intermediate FIB4 results between 1.30 and 3.25 had a 7-fold higher risk of LREs, while aHR for the intermediate VCTE results (8-12 kPa) reached borderline significance (aHR 3.9; 95% CI 0.9-18.0). FIB4 results >3.25 were found in 6.5% of patients and were associated with a 30-fold increased risk of LREs (aHR 29.5; 95% CI 10.6-82.3). VCTE results >12.0 kPa were seen in 28.1% of patients and were associated with a 21-fold increased risk of LREs (aHR 20.5; 95% CI 4.9-86.5). Fig. S4 shows the risk stratification in the

Table 1. Baseline characteristics of the study cohort.

	All (n = 1,057)	Liver biopsy available (n = 594)	No liver biopsy (n = 463)	p value
Age (years)	55.4 [45.5, 62.5]	54.5 [45.9, 62.2]	56.2 [44.9, 63.0]	0.562
Male sex (%)	659 (62.3)	365 (61.4)	294 (63.5)	0.523
Antidiabetic treatment (%)	387 (36.6)	223 (37.5)	164 (35.4)	0.480
Antihypertensive treatment (%)	469 (44.4)	256 (43.1)	213 (46.0)	0.350
Lipid-lowering treatment (%)	284 (26.9)	145 (24.4)	139 (30.0)	0.042
AST (IU/L)	40 [30, 56]	42 [31, 60]	38 [29, 50]	<0.001
ALT (IU/L)	62 [42, 88]	65 [46, 90]	59 [38, 85]	0.003
GGT (IU/L)	73 [41, 146]	81 [45, 156]	66 [38, 136]	0.004
Platelets (10 ⁹ /L)	222 [185, 262]	219 [183, 257]	224 [188, 265]	0.117
FIB4	1.27 [0.83, 1.93]	1.30 [0.86, 2.04]	1.22 [0.81, 1.86]	0.027
VCTE (kPa)	9.4 [6.9, 12.6]	10.4 [7.6, 14.0]	8.3 [6.3, 11.4]	<0.001
Biopsy length (mm)	-	20 [15, 30]	-	-
NASH (%)	-	310 (52.2)	-	-
Fibrosis stage (%):				
F0	-	166 (8.0)	-	-
F1	-	142 (3.9)	-	-
F2	-	106 (7.8)	-	-
F3	-	138 (23.2)	-	-
F4	-	42 (7.1)	-	-

Continuous variables are expressed as medians with 1st and 3rd quartiles; categorical variables are presented as numbers with percentages into bracket.

ALT, alanine aminotransferase; AST, aspartate aminotransferase; FIB4, fibrosis-4; NASH, non-alcoholic steatohepatitis; VCTE, vibration-controlled transient elastography.

^aLiver-related event: composite endpoint combining cirrhosis complications and/or hepatocellular carcinoma.

^bCirrhosis complications: ascites, spontaneous bacterial peritonitis, hepatorenal syndrome, gastrointestinal varices, variceal bleeding, liver failure, encephalopathy.

Table 2. Liver-related events during the follow-up.

	All (n = 1,057)	Liver biopsy available (n = 594)	No liver biopsy (n = 463)
Follow-up (years)	3.1 [1.3, 6.4]	2.2 [0.7, 5.6]	3.9 [2.3, 6.9]
Events during follow-up (n, %):			
Bariatric surgery	16 (1.5)	9 (1.5)	7 (1.5)
Liver-related event ^a	62 (5.9)	42 (7.1)	20 (4.3)
Cirrhosis complication ^b	52 (4.9)	38 (6.4)	14 (3.0)
Hepatocellular carcinoma	20 (1.9)	8 (1.3)	12 (2.6)
Death	32 (3.0)	12 (2.0)	20 (4.3)
Liver-related death	12 (1.1)	5 (0.8)	7 (1.5)

Continuous variables are expressed as medians with 1st and 3rd quartiles; categorical variables are presented as number with percentages in brackets.

^aLiver-related event: composite endpoint combining cirrhosis complications and/or hepatocellular carcinoma.

^bCirrhosis complications: ascites, spontaneous bacterial peritonitis, hepatorenal syndrome, gastrointestinal varices, variceal bleeding, liver failure, encephalopathy.

biopsy group according to the subgroups defined by FIB4, VCTE, and histology. Liver biopsy and VCTE showed a similar pattern, both distinguishing 2 subgroups of patients with clearly different prognoses, whereas FIB4 individualized 3 subgroups including 1 with intermediate prognosis.

Impact of diabetes on the risk of LREs

Diabetes was an independent predictor of LREs in both multi-variable analyses we performed for FIB4 and VCTE. We therefore further explored the potential impact of diabetes in patients with FIB4 <1.30, as guidelines indicate these patients should be kept at the primary care level with no need for referral to a liver specialist. In patients with baseline FIB4 <1.30, Kaplan-Meier curves showed that prognosis was impaired in patients with diabetes compared to those without (Fig. S5). Eight patients experienced LREs during follow-up despite a baseline FIB4 <1.30 (Table S1). Interestingly, 6 of these patients had diabetes and all VCTE results were increased from baseline. For patients with available FIB4 monitoring, the blood test result progressively increased, reaching FIB4 ≥1.30 before the occurrence of LREs.

The scatter plot in Fig. S6 shows the correlation between FIB4 and VCTE results at baseline. Fifty percent of the patients with baseline FIB4 <1.30 had an elevated VCTE result >8.0 kPa, and this rate was higher in patients with diabetes compared to those without (77% vs. 40%, $p < 0.001$; Fig. S7A). Fifty-five percent of patients with baseline FIB4 <1.30 had a liver biopsy available within 1 year. In this subgroup, the proportion of patients with advanced fibrosis despite FIB4 <1.30 was 8-fold higher in patients with diabetes than it was in those without: 34% vs. 4%, respectively ($p < 0.001$). Among patients with diabetes and advanced fibrosis despite baseline FIB4 <1.30, VCTE results were ≥8.0 kPa in 88% of the cases (Fig. S7B). These results suggest that a FIB4 result <1.30 should be considered with caution in patients with diabetes, and that VCTE could help to refine the evaluation in this situation.

Table 3. Comparison of Harrell C-indexes for the prediction of liver-related events.

Fibrosis test	Whole study cohort	Biopsy group
FIB4	0.817 (0.768–0.866)	0.775 (0.712–0.837)
VCTE	0.878 (0.835–0.921)	0.881 (0.832–0.931)
Biopsy	-	0.932 (0.910–0.955)
Comparison (p)		
FIB4 vs. VCTE	0.059	0.015
FIB4 vs. biopsy	-	<0.001
VCTE vs. biopsy	-	0.164

FIB4, fibrosis-4; VCTE, vibration-controlled transient elastography.

Prognostic accuracy of the FIB4-VCTE algorithm

We finally studied the prognostic significance of the stepwise FIB4-VCTE algorithm (Fig. S1). The aHRs for LREs were not significantly different between the first 2 categories of the algorithm ("FIB4 <1.30" and "FIB4 ≥1.30 then VCTE <8.0 kPa"; Table 5). These 2 categories included 63% of patients who therefore do not require referral to a liver specialist due to a good prognosis. The aHRs for LREs were significantly higher in the "Increased risk" and "High risk" categories of the FIB4-VCTE algorithm, demonstrating that these patients need specialized management (Fig. 2). The competing risk regression showed results similar to those of the main analysis (Table S2). LRE-free survival according to the 4 categories of the FIB4-VCTE algorithm is depicted in Fig. 3.

Fibrosis stages as a function of the FIB4-VCTE algorithm categories in the biopsy group are shown in Table S3. The percentages of patients correctly classified for the diagnosis of advanced fibrosis in the 3 non-invasive categories, "FIB4 <1.30", "FIB4 ≥1.30 then VCTE <8.0 kPa", and "FIB4 ≥1.30 then VCTE >12.0 kPa" were 89%, 85%, and 70%, respectively.

Discussion

NAFLD now affects 25% of the worldwide population and makes up a significant part of the burden of chronic liver diseases.²⁹ Proportionally, only a minority of patients with NAFLD develop end-stage liver disease or hepatocellular carcinoma. The prevalence of NAFLD is projected to increase even further in the coming years mainly due to the rising prevalence of obesity and metabolic syndrome.³⁰ Therefore, the challenge in current clinical practice is the accurate identification of patients at risk of developing LREs, especially considering the likely arrival of new pharmacological treatments. Thus, one key question would be whether non-invasive tests can be used instead of liver biopsy to predict which patients with NAFLD are at a higher risk of liver-related morbidity and mortality over time. Our study shows that the blood test FIB4 and VCTE stratify the risk of LREs in patients with NAFLD, and that VCTE has similar prognostic accuracy as liver biopsy. Such results suggest that these tests are acceptable alternatives to liver biopsy for prognostication and could possibly guide decisions on treatment allocation.

The optimal thresholds of non-invasive tests for the diagnosis of advanced fibrosis in NAFLD are still being debated. Indeed, depending on the sample studied, published cut-offs show different results for sensitivity, specificity, and predictive values,⁷ which brings into question their generalizability to daily clinical practice, especially in primary care. The aHRs from our

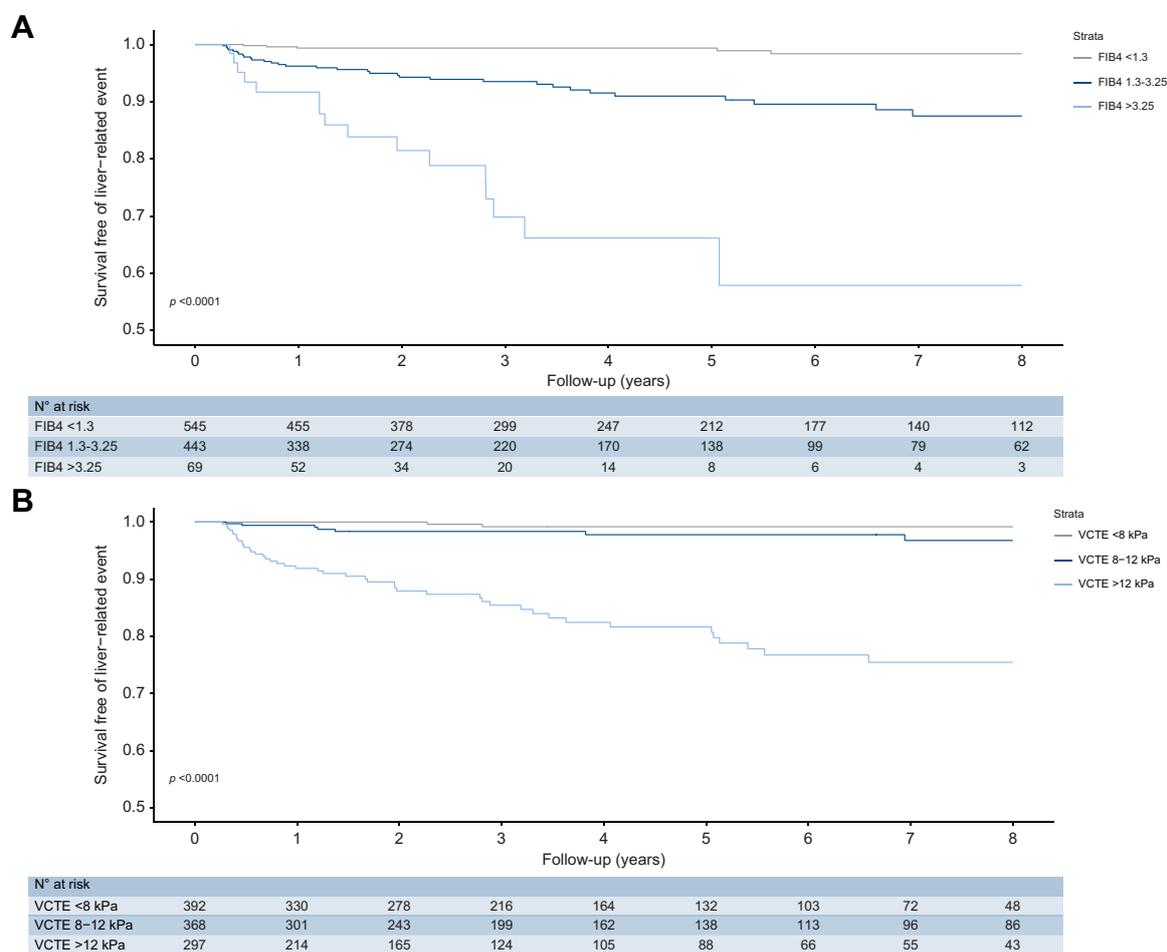


Fig. 1. Survival free of liver-related events according to the groups defined by the FIB4 or VCTE thresholds. FIB4, fibrosis-4; VCTE, vibration-controlled transient elastography.

multivariable analyses demonstrate that, beyond their diagnostic accuracy, the thresholds for FIB4 (1.30 and 3.25) and VCTE (8.0 and 12.0 kPa) delineate patient subgroups with different prognoses. In our study, the risk of LREs was very low for values below the lower cut-off, whereas that risk was somewhat

increased between the 2 cut-offs and was highest above the higher cut-off. Taken together, these results suggest that FIB4, elastography, and histology should no longer be put into competition in diagnostic studies but rather be considered as 3 different approaches (respectively functional, physical, and

Table 4. Unadjusted and adjusted hazard ratios for the prediction of liver-related events according to baseline FIB4 or VCTE results.

Fibrosis test	Stratification	Patients (n)	Events (n)	Incidence rate ^b	Univariate analysis		Multivariable analysis ^a	
					HR	[95% CI]	aHR	[95% CI]
FIB4	<1.30	545	8	2.6	1.0	(ref)	1.0	(ref)
	1.30-3.25	443	37	20.8	8.9	[3.9-20.2]	6.7	[2.7-16.5]
	>3.25	69	17	90.8	36.3	[14.8-89.5]	29.5	[10.6-82.3]
VCTE	<8.0 kPa	392	2	1.3	1.0	(ref)	1.0	(ref)
	8.0-12.0 kPa	368	11	5.2	5.0	[1.1-22.9]	3.9	[0.9-18.0]
	>12.0 kPa	297	49	37.5	32.8	[8.0-135.3]	20.5	[4.9-86.5]

(a)HR, (adjusted) hazard ratio; FIB4, fibrosis-4; VCTE, vibration-controlled transient elastography.

^aMultivariable Cox Model adjusted for age, sex, antidiabetic treatment, antihypertensive treatment, and lipid-lowering treatment.

^bPer 1,000 person/years.

Table 5. Unadjusted and adjusted hazard ratios for the prediction of liver-related events according to the groups defined by the FIB4-VCTE algorithm.

Stratification	Patients (n)	Events (n)	Incidence rate ^b	Univariate analysis		Multivariable analysis ^a	
				HR	[95% CI]	aHR	[95% CI]
FIB4 <1.30	545	8	2.6	1.0	(ref)	1.0	(ref)
FIB4 ≥1.30 & VCTE < 8.0	119	2	4.4	1.7	[0.4–8.4]	1.3	[0.3–6.8]
FIB4 ≥1.30 & VCTE 8.0–12.0	163	10	14.9	6.2	[2.3–16.3]	3.8	[1.3–10.9]
FIB4 ≥1.30 & VCTE >12.0	230	42	50.3	20.3	[9.1–45.4]	12.4	[5.1–30.2]

(a)HR, (adjusted) hazard ratio; FIB4, fibrosis-4; VCTE, vibration-controlled transient elastography.

^aMultivariable Cox Model adjusted on age, sex, antidiabetic treatment, antihypertensive treatment, and lipid-lowering treatment.

^bPer 1,000 person/years.

architectural), each able to evaluate NAFLD severity and to stratify the prognosis of the disease.

Blood tests and elastography are notable for their non-invasive nature and the advantages each brings to the table: the former can be easily prescribed by any physician, while the latter gives an immediate result during the consultation, and, when included in ultrasonography devices, additional information on liver morphology. It has been suggested to combine these advantages by using these non-invasive tests in a sequential algorithm.^{8,20} This manifests by the first line use of a blood test such as FIB4 that, when low, will identify patients with a low risk of advanced fibrosis and a favorable prognosis. Should FIB4 yield intermediate or high results, second line VCTE can be used to improve identification of advanced fibrosis.³¹ Such an approach has been shown to discriminate well between patients with and

without advanced fibrosis, while reducing the need for liver biopsy.^{8,20} Our results provide prognostic validation of this FIB4-VCTE sequential approach by showing that patients with FIB4 <1.30 and those with FIB4 ≥1.30 and VCTE <8.0 kPa have excellent prognoses, while those with FIB4 ≥1.30 and VCTE ≥8.0 kPa are at an increased risk of LREs. Our results are important because they validate the recent proposal of using the sequential FIB4-VCTE algorithm as a pathway between primary care physicians and liver specialists.²¹ Indeed, our results demonstrate that using this pathway can retain patients with good prognoses in the primary care setting, while those in need of specialized management because of an impaired prognosis can be referred to a liver specialist. This approach has been suggested to be cost-effective in the UK and the US,^{32,33} and such evaluations should now be performed in other countries.

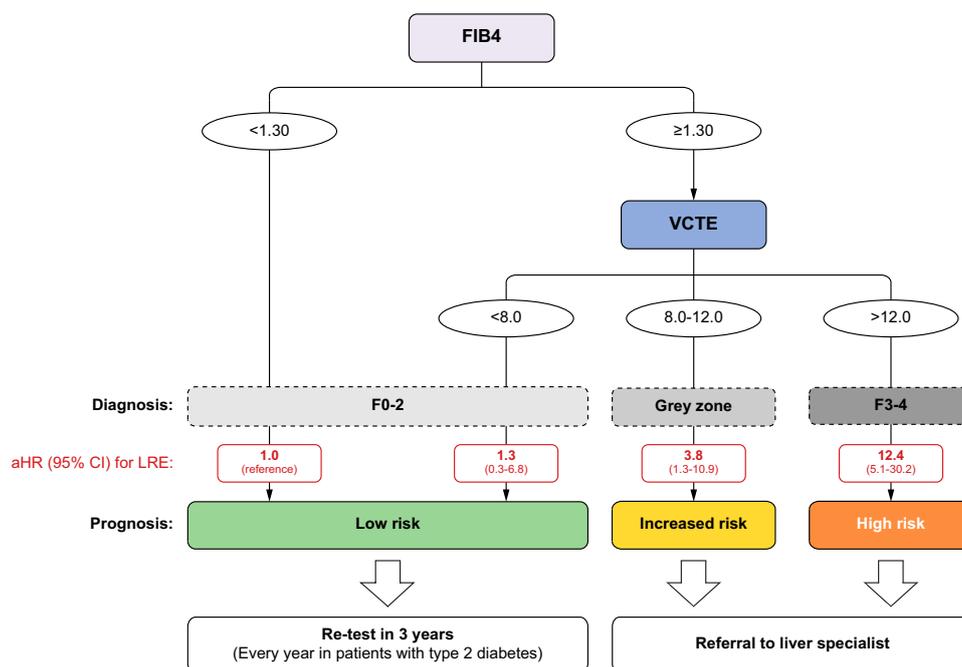


Fig. 2. Prognostic significance of the stepwise FIB4-VCTE algorithm. The stepwise FIB4-VCTE algorithm was initially developed for the non-invasive diagnosis of advanced liver fibrosis and has been recently proposed as a pathway for patient referral to the liver specialists. Our results provide prognostic validation of this pathway in NAFLD. Patients diagnosed with no/mild liver fibrosis (F0-2) according to the algorithm have a low risk of liver-related events and therefore do not need to be referred to a liver specialist. In contrast, patients in the grey zone of the algorithm and, even more so, those diagnosed with advanced fibrosis (F3-4) have an increased risk of liver-related events and therefore require specialized management of their liver disease. FIB4, fibrosis-4; VCTE, vibration-controlled transient elastography.

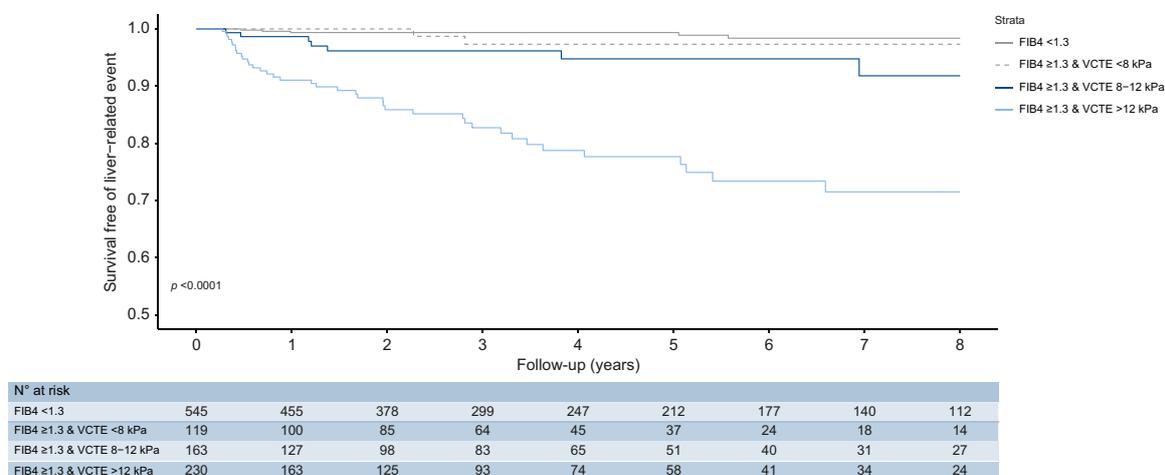


Fig. 3. Survival free of liver-related events according to the 4 groups defined by the FIB4-VCTE algorithm. FIB4, fibrosis-4; VCTE, vibration-controlled transient elastography.

Whatever the fibrosis test examined in the multivariable analysis, we consistently identified diabetes as an independent predictor of LREs. Accordingly, in the subgroup with FIB4 <1.30, patients with diabetes showed an impaired prognosis compared to those without. This is problematic, as current guidelines indicate that all patients with FIB4 <1.30 should not be referred to a liver specialist.

A first simple option to circumvent this limitation would be to repeat the FIB4 measurement during follow-up. Indeed, a second FIB4 evaluation within 5 years improves the identification of patients with increased liver-related risk in the general population.³⁴ In our work, most of the patients who experienced a LRE despite baseline FIB4 <1.30 had at least 1 additional FIB4 during their follow-up. In almost all these patients, FIB4 became ≥1.30 several years before the LRE, suggesting repeated measurements of FIB4 are useful to monitor liver-related risks in patients with NAFLD. Another option would be to perform another fibrosis test. Indeed, most of the patients with advanced fibrosis despite baseline FIB4 <1.30 had diabetes and elevated liver stiffness. This suggests that in patients with diabetes and FIB4 <1.30, VCTE may be useful in identifying the subset with impaired prognosis.

These results coming from *post hoc* analyses and a small subset of the study cohort are exploratory, so they should be considered with caution and need validation. Previous works have suggested that advanced fibrosis is associated with the duration of diabetes,³⁵ and the association with diabetes control remains debated. Finally, further works are required to better define the profile of patients with diabetes who could benefit from an additional VCTE examination despite FIB4 results <1.30.

We acknowledge some limitations in the current work. Follow-up was rather short and this limited our ability to ascertain LREs, which resulted in wide confidence intervals. However, the size of our cohorts ensured enough outcomes for a meaningful analysis. Further studies with longer follow-up duration and sample sizes conducted in real-life cohorts are required to validate our findings. As expected, patients from the biopsy group had more advanced liver disease than those

without (higher serum transaminases, higher results for non-invasive tests of liver fibrosis), reflecting the indications for liver biopsy according to international practice guidelines. Slight differences were consequently observed between the whole study cohort and the biopsy group for the FIB4 and VCTE Harrell C-indexes. The prognostic accuracies reported for FIB4, VCTE, and the FIB4-VCTE algorithm in the whole study cohort are more prone to generalization than those observed in the biopsy group, as the latter corresponds to a subset of selected patients. In a perfect study, liver biopsy would have been available in all patients, but such work is impossible given it is not ethical to perform this invasive procedure when there is no argument for significant liver disease.

In conclusion, FIB4 and VCTE allow for an accurate prediction of LREs in patients with NAFLD and could be used in place of liver biopsy for the identification of at-risk patients in need of specialized management. The sequential FIB4-VCTE algorithm accurately stratifies LRE risk and should be evaluated in real-life conditions as a referral pathway between primary care physicians and liver specialists.

Abbreviations

aHR, adjusted hazard ratio; ALT, alanine aminotransferase; AST, aspartate aminotransferase; FIB4, fibrosis-4; LREs, liver-related events; NAFLD, non-alcoholic fatty liver disease; NASH, non-alcoholic steatohepatitis; VCTE, vibration-controlled transient elastography.

Financial support

Intercept Pharmaceuticals provided grant support.

Conflict of interest

Jerome Boursier: Board advisory for Bristol-Myers, Echosens, Intercept. Research grants to his institution from EchoSens, Intercept, Inventiva, Siemens. Speaker fees from Gilead, Intercept, Lilly, Siemens. Hannes Hagström: Board advisory for Bristol-Myers Squibb and Gilead. Research grants to his

institution from Astra Zeneca, Pfizer, Gilead, MSD, EchoSens, Intercept. Mattias Ekstedt: Member of advisory board for AMRA Medical. Research grants to his institution from Intercept, Astra Zeneca and Gilead. Speaker fees from Olink. Manuel Romero-Gomez: Research grants to his institution from Intercept. Speaker and consulting fees from Intercept. Other authors have no conflict of interest to declare.

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

Authors' contributions

Study design: J. Boursier, H Hagström, M Ekstedt, S Cure, A Trylesinski, M Romero-Gomez; Data acquisition: J. Boursier, H Hagström, M Ekstedt, J Ampuero, P Nasr, L Tallab, CM Canivet, S Kechagias, Y Sánchez, E Dincuff, A Lucena, M Romero-Gomez; Analysis: J. Boursier, C Moreau, M Roux, J Riou; Drafting/ critical revision: J. Boursier, H Hagström, M Ekstedt, M Bonacci, S Cure, M Roux, A Trylesinski, M Romero-Gomez.

Data availability statement

Data can be available upon reasonable request to the Corresponding Author.

Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jhep.2021.12.031>.

References

- Taylor RS, Taylor RJ, Bayliss S, Hagstrom H, Nasr P, Schattenberg JM, et al. Association between fibrosis stage and outcomes of patients with nonalcoholic fatty liver disease: a systematic review and meta-analysis. *Gastroenterology* 2020;158:1611–1625 e1612.
- Geier A, Boursier J. Non-invasive diagnosis of patients with 'at-risk' NAFLD: only fibrosis counts? *Gut* 2020;69:1164–1165.
- EASL-EASD-EASO Clinical Practice Guidelines for the management of non-alcoholic fatty liver disease. *J Hepatol* 2016;64:1388–1402.
- Chalasanani N, Younossi Z, Lavine JE, Charlton M, Cusi K, Rinella M, et al. The diagnosis and management of nonalcoholic fatty liver disease: practice guidance from the American Association for the Study of Liver Diseases. *Hepatology* 2018;67:328–357.
- Ratziv V, Charlotte F, Heurtier A, Gombert S, Giral P, Bruckert E, et al. Sampling variability of liver biopsy in nonalcoholic fatty liver disease. *Gastroenterology* 2005;128:1898–1906.
- Davison BA, Harrison SA, Cotter G, Alkhoury N, Sanyal A, Edwards C, et al. Suboptimal reliability of liver biopsy evaluation has implications for randomized clinical trials. *J Hepatol* 2020;73:1322–1332.
- Xiao G, Zhu S, Xiao X, Yan L, Yang J, Wu G. Comparison of laboratory tests, ultrasound, or magnetic resonance elastography to detect fibrosis in patients with nonalcoholic fatty liver disease: a meta-analysis. *Hepatology* 2017;66:1486–1501.
- Boursier J, Guillaume M, Leroy V, Irls M, Roux M, Lannes A, et al. New sequential combinations of non-invasive fibrosis tests provide an accurate diagnosis of advanced fibrosis in NAFLD. *J Hepatol* 2019;71:389–396.
- Vali Y, Lee J, Boursier J, Spijker R, Loffler J, Verheij J, et al. Enhanced liver fibrosis test for the non-invasive diagnosis of fibrosis in patients with NAFLD: a systematic review and meta-analysis. *J Hepatol* 2020;73:252–262.
- Mehta SH, Lau B, Afdhal NH, Thomas DL. Exceeding the limits of liver histology markers. *J Hepatol* 2009;50:36–41.
- Angulo P, Bugianesi E, Bjornsson ES, Charatcharoenwithaya P, Mills PR, Barrera F, et al. Simple noninvasive systems predict long-term outcomes of patients with nonalcoholic fatty liver disease. *Gastroenterology* 2013;145:782–789 e784.
- Boursier J, Vergniol J, Guillet A, Hiriart JB, Lannes A, Le Bail B, et al. Diagnostic accuracy and prognostic significance of blood fibrosis tests and liver stiffness measurement by FibroScan in non-alcoholic fatty liver disease. *J Hepatol* 2016;65:570–578.
- Munteanu M, Pais R, Peta V, Deckmyn O, Moussalli J, Ngo Y, et al. Long-term prognostic value of the FibroTest in patients with non-alcoholic fatty liver disease, compared to chronic hepatitis C, B, and alcoholic liver disease. *Aliment Pharmacol Ther* 2018;48:1117–1127.
- Hagstrom H, Nasr P, Ekstedt M, Stal P, Hultcrantz R, Kechagias S. Accuracy of noninvasive scoring systems in assessing risk of death and liver-related endpoints in patients with nonalcoholic fatty liver disease. *Clin Gastroenterol Hepatol* 2019;17:1148–1156 e1144.
- Sterling RK, Lissen E, Clumeck N, Sola R, Correa MC, Montaner J, et al. Development of a simple noninvasive index to predict significant fibrosis in patients with HIV/HCV coinfection. *Hepatology* 2006;43:1317–1325.
- Srivastava A, Gailer R, Tanwar S, Trembling P, Parkes J, Rodger A, et al. Prospective evaluation of a primary care referral pathway for patients with non-alcoholic fatty liver disease. *J Hepatol* 2019;71:371–378.
- Castera L, Forns X, Alberti A. Non-invasive evaluation of liver fibrosis using transient elastography. *J Hepatol* 2008;48:835–847.
- Papatheodoridi M, Hiriart JB, Lupsor-Platon M, Bronte F, Boursier J, Elshaarawy O, et al. Refining the Baveno VI elastography criteria for the definition of compensated advanced chronic liver disease. *J Hepatol* 2021;74:1109–1116.
- Mozes FE, Lee JA, Selvaraj EA, Jayaswal ANA, Trauner M, Boursier J, et al. Diagnostic accuracy of non-invasive tests for advanced fibrosis in patients with NAFLD: an individual patient data meta-analysis. *Gut* 2021.
- Petta S, Wong VW, Camma C, Hiriart JB, Wong GL, Vergniol J, et al. Serial combination of non-invasive tools improves the diagnostic accuracy of severe liver fibrosis in patients with NAFLD. *Aliment Pharmacol Ther* 2017;46:617–627.
- Tsochatzis EA, Newsome PN. Non-alcoholic fatty liver disease and the interface between primary and secondary care. *Lancet Gastroenterol Hepatol* 2018;3:509–517.
- Kleiner DE, Brunt EM, Van Natta M, Behling C, Contos MJ, Cummings OW, et al. Design and validation of a histological scoring system for nonalcoholic fatty liver disease. *Hepatology* 2005;41:1313–1321.
- Boursier J, Anty R, Vonghia L, Moal V, Vanwolleghem T, Canivet CM, et al. Screening for therapeutic trials and treatment indication in clinical practice: MACK-3, a new blood test for the diagnosis of fibrotic NASH. *Aliment Pharmacol Ther* 2018;47:1387–1396.
- Bedossa P, Consortium FP. Utility and appropriateness of the fatty liver inhibition of progression (FLIP) algorithm and steatosis, activity, and fibrosis (SAF) score in the evaluation of biopsies of nonalcoholic fatty liver disease. *Hepatology* 2014;60:565–575.
- Stekhoven DJ, Buhlmann P. MissForest—non-parametric missing value imputation for mixed-type data. *Bioinformatics* 2012;28:112–118.
- Harrell Jr FE, Califf RM, Pryor DB, Lee KL, Rosati RA. Evaluating the yield of medical tests. *JAMA* 1982;247:2543–2546.
- Kang L, Chen W, Petrick NA, Gallas BD. Comparing two correlated C indices with right-censored survival outcome: a one-shot nonparametric approach. *Stat Med* 2015;34:685–703.
- Fine GP, Gray RJ. A proportional hazards model for the subdistribution of a competing risk. *J Am Stat Assoc* 1999;94:496–509.
- Younossi Z, Anstee QM, Marietti M, Hardy T, Henry L, Eslam M, et al. Global burden of NAFLD and NASH: trends, predictions, risk factors and prevention. *Nat Rev Gastroenterol Hepatol* 2018;15:11–20.
- Estes C, Anstee QM, Arias-Loste MT, Bantel H, Bellentani S, Caballeria J, et al. Modeling NAFLD disease burden in China, France, Germany, Italy, Japan, Spain, United Kingdom, and United States for the period 2016–2030. *J Hepatol* 2018;69:896–904.
- Castera L, Friedrich-Rust M, Loomba R. Noninvasive assessment of liver disease in patients with nonalcoholic fatty liver disease. *Gastroenterology* 2019;156:1264–1281 e1264.
- Vilar-Gomez E, Lou Z, Kong N, Vuppalanchi R, Imperiale TF, Chalasanani N. Cost effectiveness of different strategies for detecting cirrhosis in patients with nonalcoholic fatty liver disease based on United States health care system. *Clin Gastroenterol Hepatol* 2020;18:2305–2314 e2312.
- Srivastava A, Jong S, Gola A, Gailer R, Morgan S, Sennett K, et al. Cost-comparison analysis of FIB-4, ELF and fibroscan in community pathways for non-alcoholic fatty liver disease. *BMC Gastroenterol* 2019;19:122.
- Hagstrom H, Talback M, Andreasson A, Walldius G, Hammar N. Repeated FIB-4 measurements can help identify individuals at risk of severe liver disease. *J Hepatol* 2020;73:1023–1029.
- Kwok R, Choi KC, Wong GL, Zhang Y, Chan HL, Luk AO, et al. Screening diabetic patients for non-alcoholic fatty liver disease with controlled attenuation parameter and liver stiffness measurements: a prospective cohort study. *Gut* 2016;65:1359–1368.

3.2 Documents complémentaires

Non-invasive tests accurately stratify patients with NAFLD based on their risk of liver-related events

Jérôme Boursier^{1,2}, Hannes Hagström^{3,4,5}, Mattias Ekstedt⁶,
Clémence Moreau², Martin Bonacci^{7,8}, Sandrine Cure^{7,8}, Javier Ampuero⁹,
Patrik Nasr⁶, Lilian Tallab⁴, Clémence M Canivet^{1,2}, Stergios Kechagias⁶,
Yolanda Sánchez⁹, Eloise Dincuff¹, Ana Lucena⁹, Marine Roux²,
Jérémy Riou², Aldo Trylesinski^{7,8}, Manuel Romero-Gomez⁹

¹Service d'Hépatogastroentérologie et Oncologie Digestive, Centre Hospitalier
Universitaire d'Angers, Angers, France

²Laboratoire HIFIH UPRES EA3859, SFR ICAT 4208, Université d'Angers, Angers,
France

³Unit of Hepatology, Department of Upper GI, Karolinska University Hospital,
Stockholm, Sweden

⁴Department of Medicine, Huddinge, Karolinska Institutet, Stockholm, Sweden

⁵Unit of Clinical Epidemiology, Department of Medicine, Solna, Karolinska Institutet,
Stockholm, Sweden

⁶Department of Health, Medicine and Caring Sciences, Division of Diagnostics and
Specialist Medicine, Linköping University, Linköping, Sweden

⁷Intercept Pharmaceuticals, London, UK

⁸Intercept Pharmaceuticals, Inc., New York, NY, USA

⁹Digestive Diseases Department, Virgen del Rocío University Hospital, Institute of
Biomedicine of Seville, University of Seville, Seville, Spain

SUPPLEMENTARY MATERIAL

Supplementary patients and methods

Study cohorts

Patients were retrospectively included at the investigating centers, excepted for the Linköping cohort (prospective inclusion). Data collected included clinical, biological, and elastography data at baseline and follow-up visits, as well as clinical events at baseline and during followup.

Angers cohort

Adult patients with NAFLD seen as outpatients between 2004 and 2016 were retrospectively screened in the outpatient registry of the Hepato-Gastroenterology Department of Angers University Hospital, Angers, France. To be included in the longitudinal cohort, patients had to have a liver fibrosis evaluation (FIB4 and/or Fibroscan examination and/or liver biopsy), and to have attended at least two additional follow-up visits in our Department. At Angers center, the choice between the M and the XL probe of the Fibroscan was done according to the BMI: M probe for the patients with BMI <30 kg/m² and XL probe for those with BMI ≥30 kg/m² [1].

Spanish cohort

Adult patients with NAFLD were retrospectively screened between 2010 and 2017 in the Pathology Department of Virgen del Rocío University Hospital and Valme University Hospital, Seville, Spain. To be included in the longitudinal cohort, patients had to have a liver fibrosis evaluation (FIB4 and/or Fibroscan examination and/or liver biopsy), and to have attended at least three visits (baseline and 2 follow-up) in the UCM of Digestive Diseases in both hospitals. These patients were also included in the Hepamet Spanish Registry of NAFLD/NASH [2]. Fibroscan examination started with the M probe; if results were of low quality or the automated software so recommended, a switch to the XL probe was done.

Stockholm cohort

Adult patients with unequivocal NAFLD seen during 1997 and 2017 were retrospectively included from the Hepatology Departments of Karolinska University Hospital, Stockholm, Sweden and Linköping University, Linköping, Sweden. The basis for a NAFLD diagnosis was by exclusion of competing liver diseases, and either a liver biopsy or a Fibroscan examination, or both. Only patients with repeat FIB-4, Fibroscan examination, or liver biopsy were included in this study. FIB-4 measurements could be done at external sites, such as in primary care. Fibroscan examination started with the M probe; if results were of low quality or the automated software so recommended, a switch to the XL probe was done.

Linköping cohort

The Linköping cohort constitutes of adult patients prospectively included at the Department of Gastroenterology and Hepatology, University Hospital, Linköping, Sweden, for evaluation of chronically (>6 months) elevated serum alanine aminotransferase, and/or serum aspartate aminotransferase, and/or serum alkaline phosphatase [3]. At inclusion, diagnostic workup was performed, including physical examination, laboratory investigation and liver biopsy. For the present study, NALFD patients with repeat FIB-4, Fibroscan examination, or liver biopsy were included. All patients were prospectively followed-up at the Department of Gastroenterology and Hepatology in Linköping. Fibroscan examination started with the M probe; if results were of low quality or the automated software so recommended, a switch to the XL probe was done.

Supplementary Figures

Fig. S1: Stepwise FIB4-VCTE algorithm

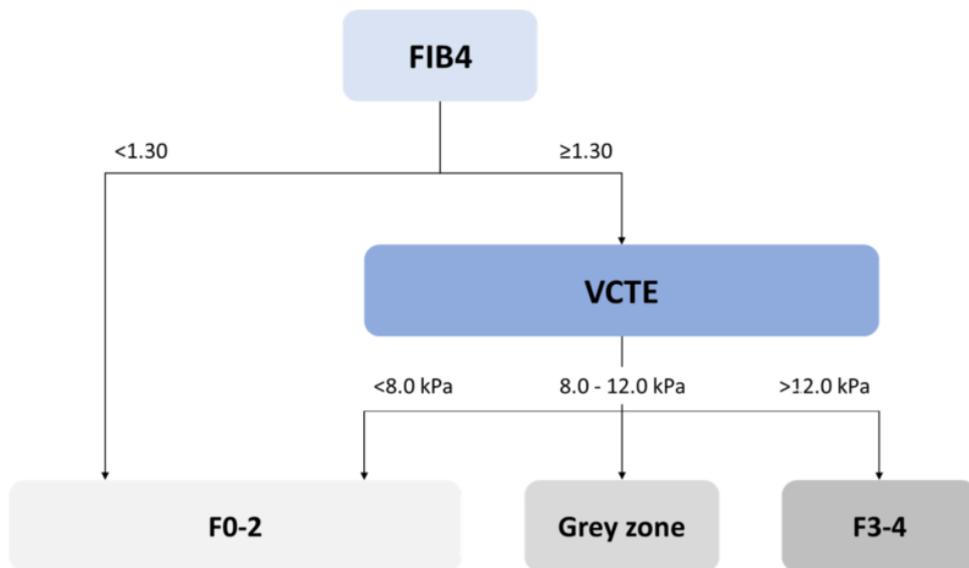


Fig. S2: Survival free of liver-related events with the Kaplan Meier method and with the Fine & Gray method considering death as competing risk

Survival curves with both methods were perfectly superimposed, showing that death as competing risk did not influence the results in our study.

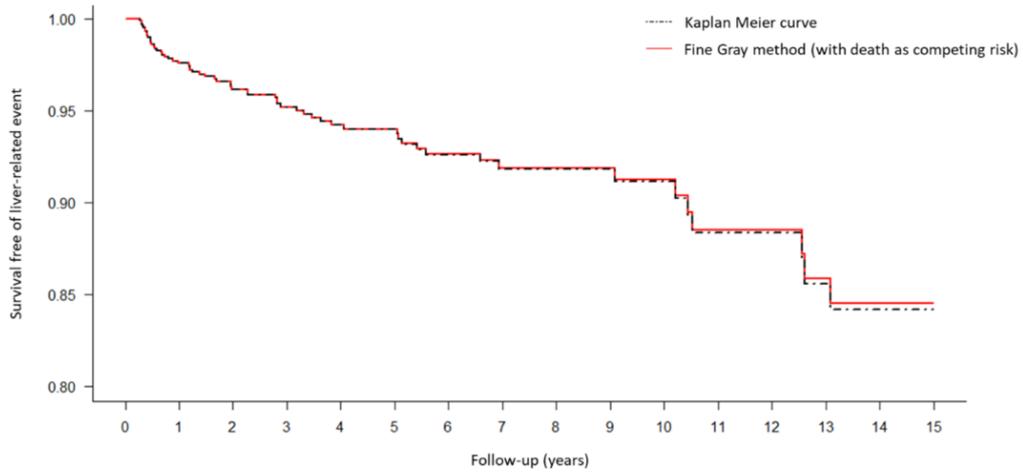


Fig. S3: Study flow chart

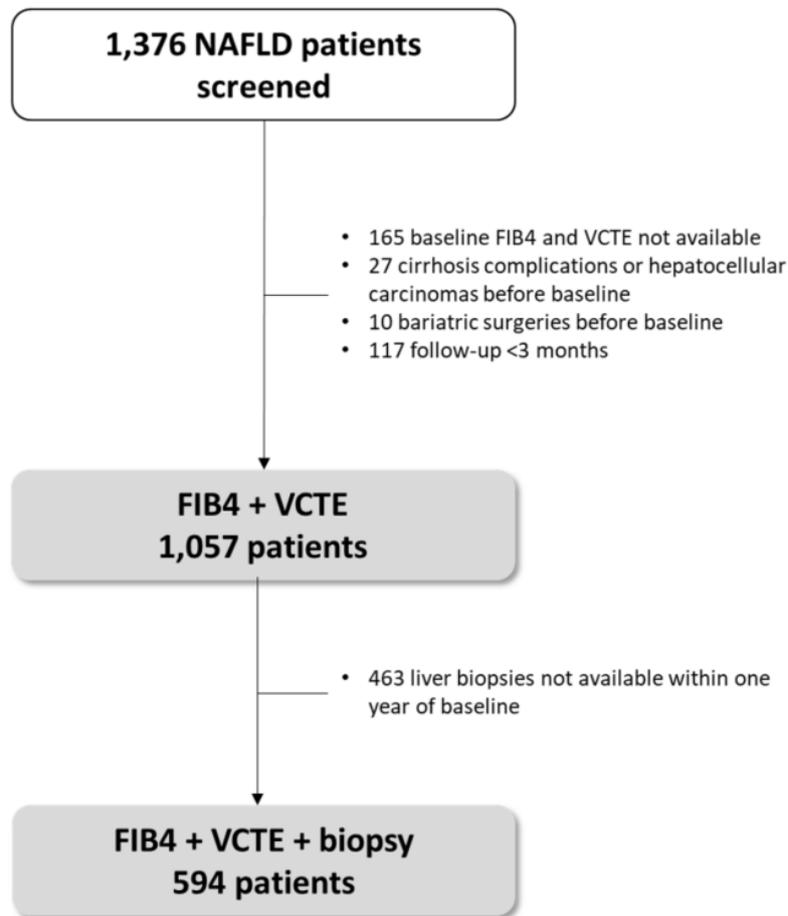
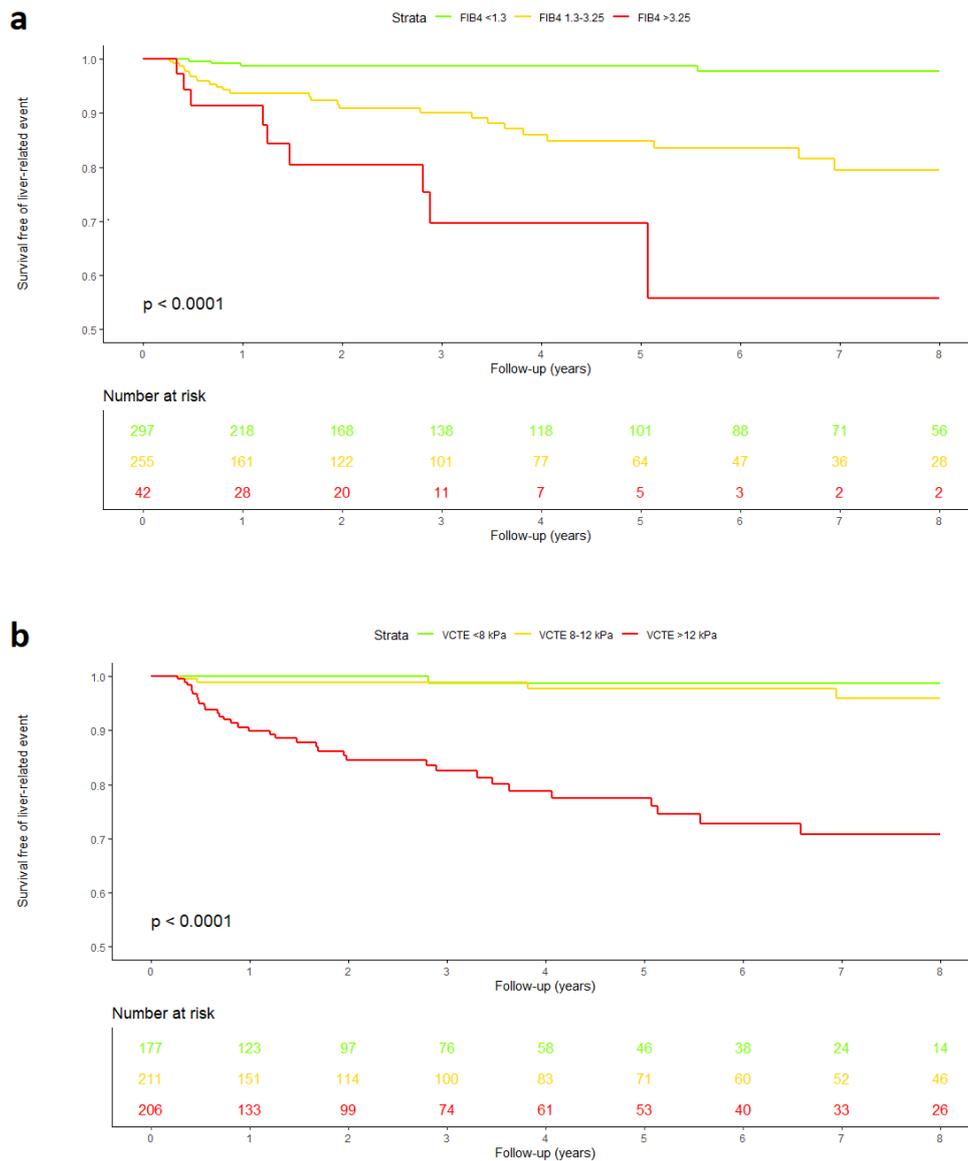


Fig. S4: Survival free of liver-related events according to baseline results for FIB4 (panel a), VCTE (panel b), and liver biopsy (panel c) in the biopsy group

Non-invasive fibrosis tests were categorized into three subgroups (low, intermediate, high risk) according to published cut-offs: 1.30 and 3.25 for FIB4, 8.0 and 12.0 kPa for VCTE.



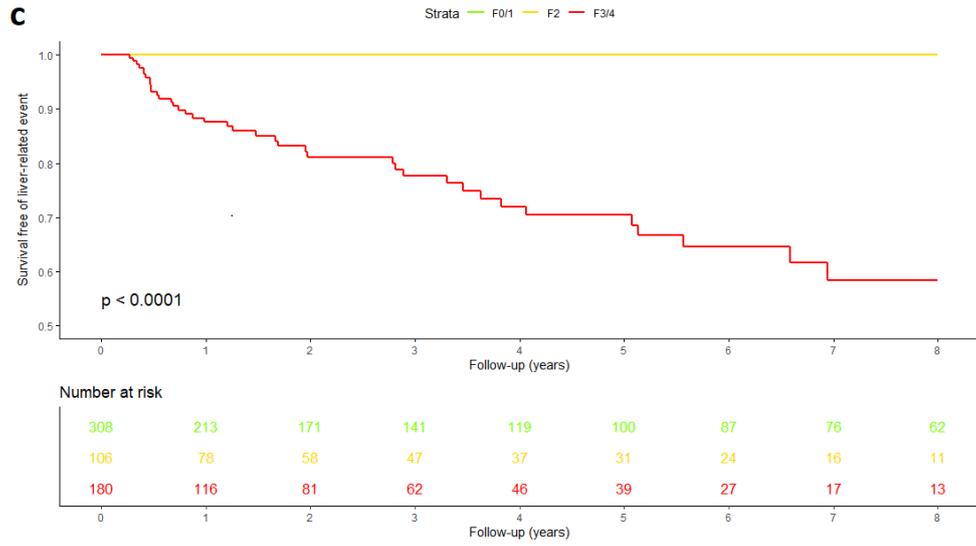


Fig. S5: Survival free of liver-related events according to diabetes status in patients with baseline FIB4 <1.30

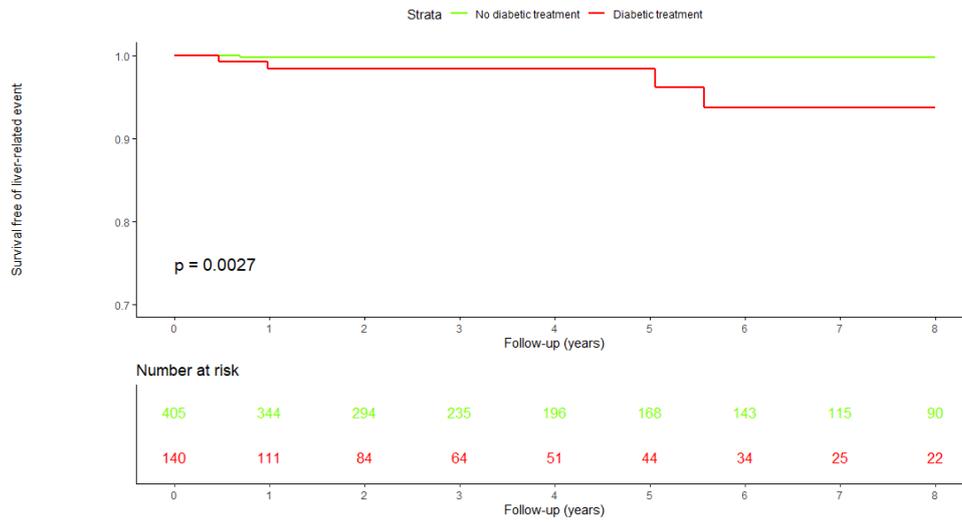


Fig. S6: Correlation between baseline FIB4 and VCTE results

aR2: 0.19 in all patients, 0.19 in patients with diabetes, 0.16 in patients without diabetes

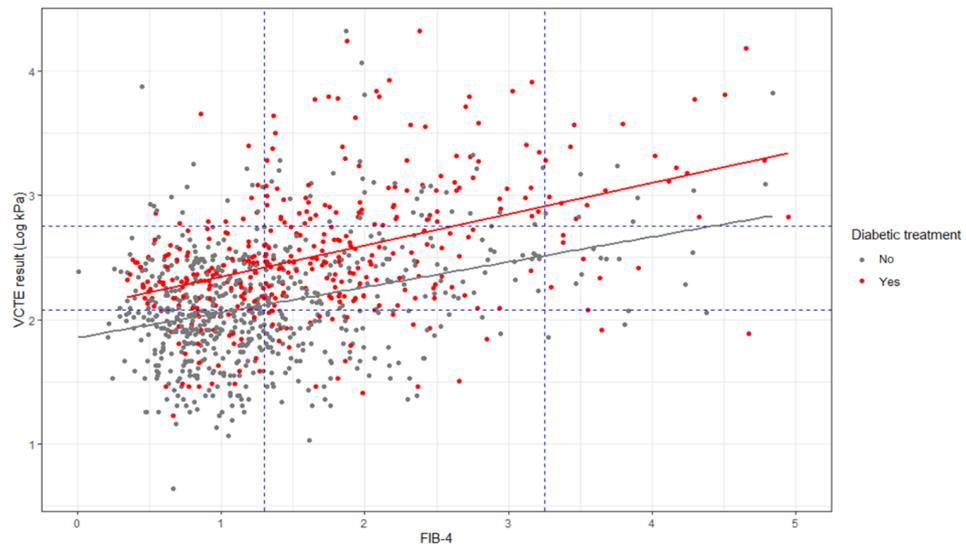
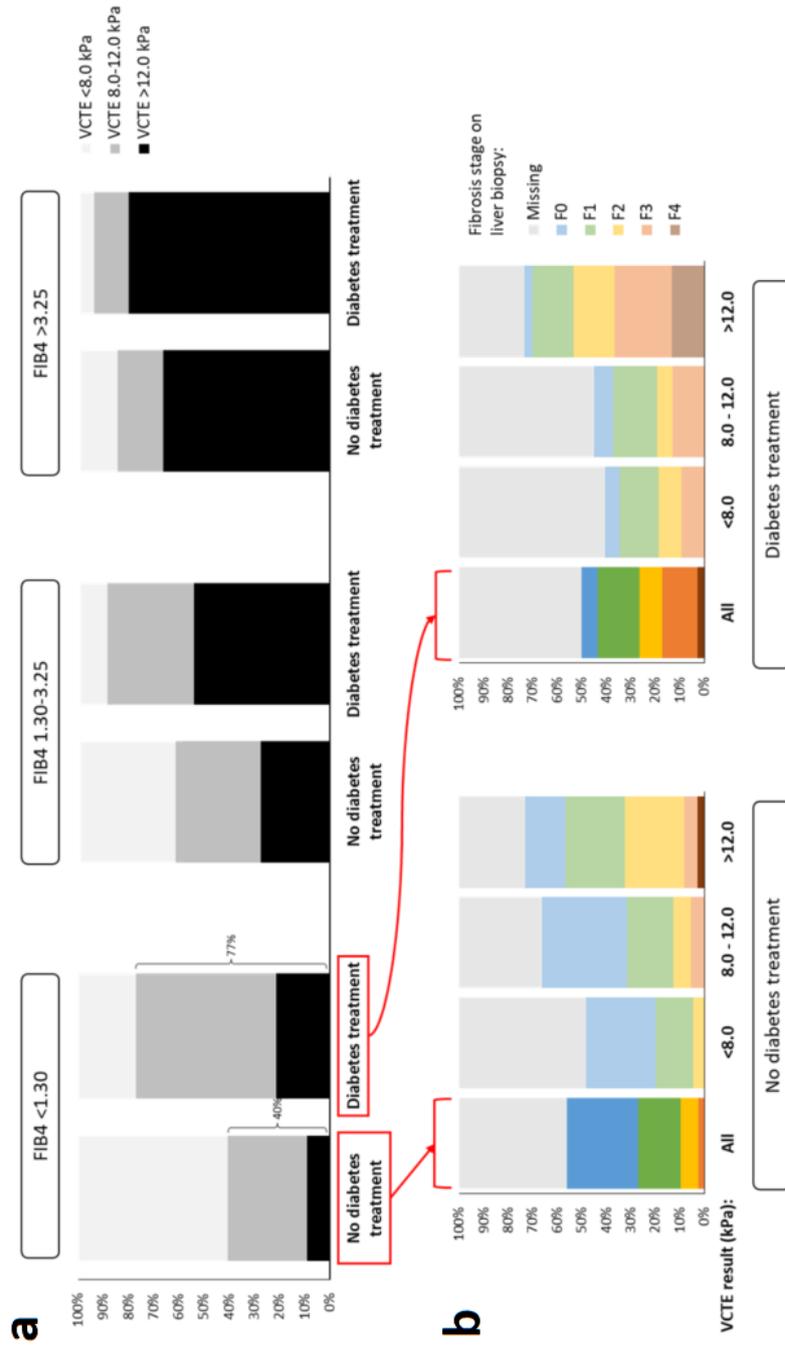


Fig. S7: VCTE result as a function of FIB4 and diabetes at baseline (panel a), and fibrosis stage in patients with baseline FIB4 <1.30 as a function of diabetes and VCTE result (panel b)



Supplementary Tables

Table S1: Follow-up of the patients who experienced liver-related events despite baseline FIB4 <1.30

Patient	Liver-related event	Visit	AST (IU/L)	ALT (IU/L)	Platelets (Giga/L)	FIB4	VCTE (kPa)	Antidiabetic treatment	Delay until event (years)
#1	Complication of cirrhosis	Baseline	48	83	279	1.194	13.3	Yes	5.1
		Follow-up #1	40	71	293	1.040	16.3	Yes	4.1
		Follow-up #2	53	76	252	1.573	19.2	Yes	3.1
		Follow-up #3	44	72	266	1.291	14.0	Yes	2.1
		Follow-up #4	55	102	257	1.423	17.3	Yes	1.1
		Follow-up #5	51	97	218	1.608	-	Yes	0.6
#2	Complication of cirrhosis	Follow-up #6	37	76	247	1.171	-	Yes	0.1
		Baseline	37	122	400	0.484	10.0	Yes	10.5
		Follow-up #1	41	104	339	0.696	-	Yes	9.6
		Follow-up #2	86	135	309	1.452	-	Yes	7.6
		Follow-up #3	94	133	322	1.559	23.3	Yes	6.7
		Follow-up #4	82	143	334	1.273	-	Yes	6.3
#3	Complication of cirrhosis	Follow-up #5	79	93	306	1.760	34.3	Yes	2.5
		Follow-up #6	70	75	320	1.719	-	Yes	0.2
		Follow-up #7	142	83	255	4.172	-	Yes	0.003
		Baseline	26	51	202	1.045	14.8	Yes	5.6
		Follow-up #1	37	81	202	1.184	-	Yes	5.4
		Follow-up #2	34	47	155	1.867	-	Yes	5.2
#4	Complication of cirrhosis	Follow-up #3	35	88	173	1.278	16.3	Yes	4.3
		Follow-up #4	55	84	158	2.270	22.6	Yes	3.8
		Follow-up #5	35	38	193	1.777	-	Yes	3.2
		Follow-up #6	37	41	180	1.956	18.8	Yes	2.6
		Follow-up #7	32	39	214	1.484	16.0	Yes	1.6
		Follow-up #8	38	42	160	2.310	21.5	Yes	0.5
#5	Complication of cirrhosis	Baseline	46	88	262	0.529	18.4	No	23.6
		Follow-up #1	50	129	247	0.737	-	No	10.5
#6	Complication of cirrhosis	Baseline	66	173	304	0.506	18.7	No	29.4
		Follow-up #1	84	90	83	6.387	-	Yes	0.2
#7	Complication of cirrhosis	Baseline	83	112	355	1.187	69.6	Yes	0.5
		Follow-up #1	26	56	249	0.709	11.8	Yes	1.0
#8	Complication of cirrhosis	Baseline	56	103	224	1.285	12.2	No	0.7
		Follow-up #1	56	103	224	1.285	12.2	No	0.7

Table S2: Subdistribution Hazard Ratios (HR) for the prediction of liver-related events according to the groups defined by the stepwise FIB4-VCTE algorithm, using competing risk regression (Fine-Gray Model)

Fibrosis test	Stratification	Patients		Events (n)	HR	[95% CI]	p-value
		(n)	(n)				
FIB4	<1.30	545	8	1.0	-	(ref)	
	1.30 - 3.25	443	37	6.2	[2.4-15.6]	<0.001	
	>3.25	69	17	27.8	[10.2-76.2]	<0.001	
VCTE	<8.0 kPa	392	2	1.0	-	(ref)	
	8.0 - 12.0 kPa	368	11	3.9	[0.9-17.6]	0.076	
	>12.0 kPa	297	49	21.2	[4.9-92.6]	<0.001	
FIB4-VCTE algorithm	FIB4 <1.3	545	8	1.0	-	(ref)	
	FIB4 ≥1.3 & VCTE <8.0	119	2	1.2	[0.3-5.8]	0.720	
	FIB4 ≥1.3 & VCTE 8.0-12.0	163	10	3.5	[1.2-9.8]	0.018	
	FIB4 ≥1.3 & VCTE >12.0	230	42	11.8	[4.7-29.4]	<0.001	

^aMultivariable model adjusted for age, sex, antidiabetic treatment, antihypertensive treatment, and lipid-lowering treatment.

Table S3: Histological fibrosis stages as a function of the FIB4-VCTE algorithm categories in the biopsy group

FIB4		≥ 1.30			
		< 1.30	< 8.0	8.0-12.0	> 12.0
VCTE (kPa)		-			
Histological fibrosis stage	F0	127	19	14	6
	F1	94	16	19	13
	F2	42	6	30	28
	F3	29	6	27	76
	F4	5	1	2	34

References

- [1] Wong VWS, Irlles M, Wong GLH, et al. Unified interpretation of liver stiffness measurement by M and XL probes in non-alcoholic fatty liver disease. *Gut* 2019; 68(11): 2057–2064.
- [2] Ampuero J, Aller R, Gallego-Durán R, et al. The effects of metabolic status on non-alcoholic fatty liver disease-related outcomes, beyond the presence of obesity. *Alimentary pharmacology & therapeutics* 2018; 48(11-12): 1260–1270.
- [3] Nasr P, Ignatova S, Kechagias S, Ekstedt M. Natural history of nonalcoholic fatty liver disease: A prospective follow-up study with serial biopsies. *Hepatology communications* 2018; 2(2): 199–210.

PRÉDICTION PERSONNALISÉE DU RISQUE DE COMPLICATIONS HÉPATIQUES APRÈS UNE RVS CHEZ LES PATIENTS INFECTÉS PAR LE VHC

Aujourd’hui, plus de 95% des patients traités contre le Virus de l’Hépatite C (VHC) guérissent grâce à la nouvelle génération de médicaments antiviraux directs. Même si cela est associé à un risque moindre de complications hépatiques, un risque résiduel subsiste chez les patients cirrhotiques. Dans ce chapitre, nous avons développé un modèle conjoint à classes latentes permettant d’obtenir des prédictions de risque personnalisées après une RVS (Réponse Virologique Soutenue). Les résultats obtenus sont en cours de soumission.

4.1 Article scientifique

Résumé L’éradication du VHC améliore significativement le pronostic des patients, mais ne supprime pas totalement le risque de décompensation de cirrhose et de Carcinome Hépatocellulaire (CHC). Par conséquent, les directives internationales recommandent de maintenir une surveillance semestrielle du CHC chez les patients cirrhotiques, indépendamment de la RVS. Cependant, en raison de la faible incidence du CHC après une RVS, une récente étude coût-efficacité a conclu que l’identification des groupes à haut risque, qui sont plus susceptibles de bénéficier d’une surveillance, serait d’un grand intérêt pour adapter les directives. D’abord, notre étude a confirmé le faible taux de complications hépatiques après la RVS. Ensuite, nous avons montré que le risque de complications post-RVS était associé au sexe du patient, ainsi qu’à la dynamique du test sanguin FIB-4 et du statut diabétique. En utilisant des statistiques innovantes, nous avons obtenu des prédictions personnalisées du risque résiduel de complications hépatiques après la RVS.

Dynamic Personalized Prediction of the Individual Liver-Related Risk after Sustained Viral Response in HCV patients

Clémence Moreau¹, Marine Roux¹, Jérémie Riou^{2,3}, Pierre Nahon^{4,5},
Fabrice Carrat^{6,7}, Jérôme Boursier^{1,8}

¹Laboratoire HIFIH, UPRES EA3859, SFR 4208, Université d'Angers, Angers, France

²Centre Hospitalier Universitaire d'Angers, Délégation à la recherche clinique et à l'innovation, Angers, France

³Inserm, UMR 1066, Micro et Nanomédecines translationnelles, Université d'Angers, Angers, France

⁴AP-HP, Hôpital Jean Verdier, Unité d'Hépatologie, Bondy, France

⁵Inserm, UMR 1162, Génomique fonctionnelle des tumeurs solides, F-75000, Paris, France

⁶AP-HP, Hôpital Saint-Antoine, Unité de Santé Publique, Paris, France

⁷Inserm, UMR-S 1136, Institut Pierre-Louis d'Epidémiologie et de Santé Publique, Paris, France

⁸Centre Hospitalier Universitaire d'Angers, Service d'Hépto-Gastroentérologie et Oncologie Digestive, Angers, France

Summary HCV eradication significantly improves patient prognosis, but does not totally eliminate the risk of decompensation of cirrhosis and Hepatocellular Carcinoma Hepatocellular Carcinoma (HCC). Therefore, international guidelines recommend maintaining biannual HCC surveillance in cirrhotic patients, independent of Sustained Viral Response (SVR). However, due to the low incidence of HCC after SVR, a recent cost-effectiveness study concluded that identifying high-risk groups, which are more likely to benefit from surveillance, would be of great interest to adapt the guidelines. First, our study confirmed the low rate of liver complications after SVR. Second, we showed that the risk of post-SVR complications was associated with patient gender, as well as with the dynamics of FIB-4 blood test and diabetic status. Using innovative statistics, we obtained personalized predictions of the residual risk of liver complications after SVR.

Keywords Cirrhosis, Hepatitis C, Hepatocellular carcinoma, Personalized medicine.

The Tables and Figures cited in this article can be found at the end of this article.

Abstract

Background Sustained Viral Response (SVR) significantly improves the prognosis in patients with Hepatitis C Virus (HCV) chronic infection, but does not totally alleviate the risk of Liver-Related Complications (LRC). We aimed to evaluate whether the dynamics of multiple measurements of simple parameters after SVR enable the development of a personalized prediction of prognosis in HCV patients.

Methods HCV mono-infected patients who experienced SVR in two prospective cohorts (ANRS CO12 CirVir cohort: derivation set; ANRS CO22 HEPATHER cohort: validation set) were included in the study. The study outcome was LRC, a composite criterion including decompensation of cirrhosis and/or hepatocellular carcinoma. Joint latent class modelling accounting for both biomarker trajectory and event occurrence during follow-up was developed in the derivation set to compute individual dynamic predictions, with further evaluation in the validation set. Calibration was assessed using the time-dependent Brier Score.

Results In the derivation set, FIB4 was identified as a biomarker associated with LRC occurrence after SVR. Using gender with the evolution of FIB4 and diabetes status, joint modelling identified three patient profiles with different probability of LRCs. In the validation set, individual dynamic predictions from the joint model accurately stratified the risk of LRCs. Time-dependent Brier Score showed good calibration that improved with the accumulation of visits, justifying our modelling approach considering both baseline and follow-up measurements.

Conclusion Multiple repeated measurements of simple parameters predict the individual residual risk of LRC after SVR. Our dynamic modelling improves personalized medicine after virus eradication in HCV patients.

1 Introduction

The management of Hepatitis C Virus (HCV) chronic infection has dramatically changed thanks to the latest generation of direct-acting antivirals that eliminate viral infection in >95% of patients [1]. Achieving Sustained Virological Response (SVR) is associated with reduced risks for cirrhosis decompensation, HepatoCellular Carcinoma (HCC), liver-related death, and all-cause mortality [2, 3, 4]. Thanks to these considerable improvements to patient prognoses, the risk of Liver-Related Complications (LRCs) persists after SVR in only a small subset of patients, especially those having cirrhosis [2, 5, 6]. After SVR, the critical challenge for physicians is to accurately identify this small subset of at-risk patients in whom specialized management must be maintained to prevent the development of cirrhosis complications or to detect HCC at an early curative stage.

Several works have developed prediction models to evaluate the liver-related prognosis after SVR in HCV patients [7]. Thanks to simple parameters measured at a single time

point (clinical data, usual biology, elastography) and combined according to different methods (survival model, deep learning), these models discriminate patient subgroups having different prognoses [7]. Another approach was to evaluate the prognostic significance of the dynamics of non-invasive fibrosis tests measured before antiviral treatment and after SVR [6, 8, 9]. Despite they provide significant advancement in the management of HCV patients who achieved SVR, these works have some limitations. Indeed, models were often developed and validated in retrospective datasets with inherent bias for patient selection and data collection. Moreover, models mostly focused on HCC prediction and not cirrhosis decompensation, the latter being also an important hard outcome in HCV patients. Many works considered only baseline parameters, some considered their evolution between two time points, but very few have assessed the prognostic significance of the dynamic profile of biomarkers evolution at multiple follow-up time points after SVR. Finally, the most important limitation is that published models aim at the stratification of patients in few subgroups, which is not in the scope of precision medicine that aims to provide a personalized prediction at the individual level.

Therefore, we aimed to develop a new personalized prediction of the individual liver-related risk after SVR in HCV patients, based on innovative statistical methods that consider the long-term dynamics of biomarkers, with validation of the prognostic model in a large prospective « real life » cohort including more than seven thousand patients.

2 Patients and methods

Patients

The present work was an ancillary study derived from the prospective CirVir and HEPATHER cohorts, both sponsored and funded by the French National Research Agency (ANRS). The ANRS CO12 CirVir cohort has included patients with compensated biopsy-proven cirrhosis due to viral B or C hepatitis from March 2006 until June 2012 in 35 French healthcare centers specialized in Hepatology [10]. The ANRS CO22 HEPATHER cohort is a French national, multicenter, prospective, observational cohort study of patients with active or inactive hepatitis B virus or past or present HCV infection, which started in August 2012 [2]. HEPATHER is a « real life » cohort that included consecutive patients in 32 expert hepatology centers. The protocol of CirVir and HEPATHER received approval from ethics committee (CirVir: *Comité de Protection des Personnes* (CPP), Aulnay-sous-Bois, France; HEPATHER: *CPP Ile de France 3*, Paris, France) and complied with the ethics guidelines of the 1975 Declaration of Helsinki. All patients provided written informed consent to participate in the cohorts. The full protocols of CirVir and HEPATHER are available via the ANRS website (<http://anrs.fr>).

For the present study, we included the HCV mono-infected patients from the CirVir and the HEPATHER cohorts who achieved SVR. We excluded those patients who had any of the following condition: history of LRCs before SVR, HCC occurrence within the six months following SVR, FIB4 measurement available at any follow-up visit, and no follow-up visit recorded.

Follow-up and study outcome

Baseline in the CirVir cohort was defined as the date of SVR or the date of inclusion in the cohort when SVR was achieved upstream of it. Patients were seen by physicians every six months as per cirrhosis monitoring recommendations, with collection of clinical, biological and ultrasonography data at each visit. For the HEPATHER cohort, baseline was defined as the date of SVR. Patients were seen by physicians every year with collection of data from the clinical practice. More information about data collection during the follow-up in CirVir and HEPATHER cohorts is provided in **Supplementary Material**. The study outcome for the present work was the occurrence of LRCs, a composite endpoint including HCC and/or cirrhosis decompensation (ascites, encephalopathy, variceal bleeding, jaundice). Follow-up was stopped at the first LRC, the patient's death, or the last follow-up visit.

Data recorded and blood fibrosis tests

Age, gender, body mass index, diabetes status, arterial hypertension and dyslipidemia were extracted from the databases of CirVir and HEPATHER cohorts. The considered biological data included liver enzymes (AST, ALT, and GGT), prothrombin time, bilirubin, albumin, and platelet count. The blood fibrosis tests APRI [11] and FIB4 [12] were calculated according to their published formulae:

$$\text{APRI} = \frac{\text{AST (ULN)} \times 100}{\text{platelets (G/L)}}$$

and

$$\text{FIB4} = \frac{\text{age} \times \text{AST (IU/L)}}{\text{platelets (G/L)} \times \text{ALT}^{\frac{1}{2}} \text{(IU/L)}}.$$

APRI and FIB4 were calculated at baseline and at each follow-up visit.

Statistical analyses

Continuous variables were expressed as medians with 1st and 3rd quartiles, and categorical data were expressed as frequencies (%). The prospective CirVir cohort was used as the derivation set for the development of the new model for personalized prediction of LRC, and the large HEPATHER cohort as the « real life » prospective validation set. We first identified the best biomarker for the prediction of LRC in the derivation set using time-dependent ROC curves [13]. The most accurate biomarker was then used to develop a prediction model with the joint latent class modelling approach [14]. The interest of such modelling is to consider in the risk equation all the repeated measurements of a biomarker during the follow-up. The individual estimated risk of LRC (given as percentage at different time horizons) is readjusted at each visit according to the whole history of the biomarker evolution, therefore providing personalized prediction. More detailed about the methods used is presented in **Supplementary Material**. Calibration of the study model was evaluated using the time-dependent Brier Score [15, 16]. Statistical analyses were performed using R version 3.5.2.

Web-tool for individualized prediction of LRCs in clinical practice

We built a Shiny web application for the clinical use of our prediction model. Using data collected during patient follow-up, this web application indicates the LRC risk at one, three and five years for any given patient.

3 Results

Patient characteristics

Derivation set Of the 1,323 HCV cirrhotic patients in the CirVir cohort, we excluded 628 patients because of no documented SVR (n=540), history of LRC before SVR (n=65), HCC within six months after SVR (n=3), and no follow-up visits recorded (n=20; **Figure 1a**). The derivation set for the study thus included 695 patients whose baseline characteristics are summarized in **Table 1**. The median follow-up was 3.8 (1.6-7.5) years during which the patients had a median of 6 (4-11) visits.

Validation set Of the 10,412 patients in the HEPATHER cohort, we excluded 3,348 patients because of no SVR achieved (n=786), no FIB4 measurement available after SVR (n=1,178), history of LRC before SVR (n=1,142), HCC occurrence within six months after SVR (n=28), no follow-up visit recorded (n=214; **Figure 1b**). The validation set for the study thus included 7,064 patients whose characteristics are summarized in **Table 1**. The median follow-up was 3.6 (2.5-4.9) years during which the patients had a median of 5 (3-6) visits.

LRC incidence

In the derivation set, 50 (7.2%) patients developed at least one LRC during the follow-up (**Table 1**). The first occurrence of LRC was HCC in 25 patients and cirrhosis decompensation in the 25 others. In the validation set, 273 (3.9%) patients had at least one LRC: 204 patients had HCC first, while 69 had cirrhosis decompensation first. Patients who developed LRCs during follow-up had more severe liver disease at baseline compared to those who did not, as shown by higher results for bilirubin, FIB4 and APRI, and lower results for platelets count, albumin, and prothrombin time (**Supplementary Tables S1 and S2**). Survival free of LRCs was 99.1% at 1 year, 96.6% at 3 years and 92.8% at 5 years in the derivation set; and 99.7% at 1 year, 97.1% at 3 years, and 94.7% at 5 years in the validation set (**Figure 2**).

Development of the study prediction model (derivation set)

In the derivation set, time-dependent ROC curves showed that blood fibrosis tests combining several biomarkers (FIB4, APRI) provided better discriminative ability than any single biomarker (AST, ALT, platelets; **Figure 3**). The time-dependent ROC curve of FIB4 was higher than that of APRI, especially at short-term and long term-follow-up, with values around 0.75 at all horizons. FIB4 was therefore chosen as the candidate biomarker for further analyses.

The Joint Latent Class Model (JLCM) used gender and repeated measurement of FIB4 and diabetes status to identify three different profiles of patients (i.e., three joint latent classes). More detailed information about the model development is presented in **Supplementary Results**. Despite preserved liver function at baseline (**Supplementary Table S4**), the prognosis was highly different between these three patient groups in the derivation set (**Supplementary Figure S1**), showing our model was able to discriminate different prognoses through gender and the evolution of simple parameters (FIB4 and diabetes) during the follow-up.

Validation of the study prediction model (validation set)

The predicted risk of LRC estimated by our model at different time horizons (one year, two years, and three years) was calculated for each patient at different visits (V0 baseline visit, V1 one-year follow-up visit, and V2 two-year follow-up visit). Cumulative incidence curves of observed LRC showed that the stratification of patients according to the risk predicted by the study model discriminated subgroups with different prognoses (**Figure 4**). **Table 2** shows the individual prediction by the model fitted well with the observed cumulative incidence of LRC at the different time horizons during the patient follow-up. As an example, for patients whose predicted risk assessed at V0 was $<1.5\%$ /year, the observed incidence of LRC was 0.4% one year after (predicted to be $<1.5\%$), 1.0% two years after (predicted to be $<3.0\%$), and 1.7% three years after (predicted to be $<4.5\%$). For patients whose predicted risk assessed at V2 was $\geq 3.0\%$ /year, the observed incidence of LRC was 5.0% one year after (predicted to be $\geq 3.0\%$), 12.5% two years after (predicted to be $\geq 6.0\%$), and 19.6% three years after (predicted to be $\geq 9.0\%$).

The low Brier Scores for the prediction of liver-related complications at the different visits and different time horizons confirmed the good calibration of our model (**Table 3**). In addition, this Brier score analysis demonstrated the benefit of accumulating visits over time for the prediction of LRC. As an example, the prediction of the three-year risk of LRC can be performed at three time points: three-year prediction at V0, two-year prediction at V1, or one-year prediction at V2. **Table 3** shows the Brier Score was higher (meaning worse prediction) for the three-year prediction at V0 as compared to the two-year prediction at V1 (0.023 vs 0.020), and the lowest (0.013) for the one-year prediction at V2. This was even more pronounced in the subgroup with the highest risk ($\geq 3.0\%$ /year) with Brier Scores at 0.131, 0.117, and 0.039 for respectively three-year prediction at V0, two-year prediction at V1, and one-year prediction at V2.

Web-based LRC risk estimating tool

We built a Shiny web application to implement our study prediction model for use in clinical practice (available at <https://hifih.shinyapps.io/lrcrisk/>). **Figure 5** provides an example of this application. It depicts a patient whose FIB4 values first decrease after SVR, then increase over time because the patient developed a metabolic syndrome and then type 2 diabetes. The associated probabilities of LRCs are adjusted over time according to the trajectory of FIB4.

4 Discussion

International guidelines recommend antiviral therapy for all adult patients with chronic HCV infection, and to maintain follow-up after SVR in cirrhotic patients [17, 18]. With the advent of non-invasive tests and highly effective treatments, physicians do no longer perform liver biopsy to their patients having chronic hepatitis C. On the other hand, non-invasive fibrosis tests have moderate positive predictive value for cirrhosis before antiviral therapy [19, 20], and the significance of their results after SVR is difficult to interpret [21, 22]. For all these reasons, the recommendations for patient monitoring based on fibrosis stage are challenging to apply in clinical practice. A more pragmatic approach is now to shift away from the reasoning based on fibrosis stage, and to design a patient management tailored to the individual residual risk of liver-related complication after SVR [23, 24]. Our study, performed in the prospective ANRS CO12 CirVir and ANRS CO22 HEPATHER cohorts, confirmed the low rate of LRCs after SVR. We chose to first develop our prediction model in the CirVir cohort where the prevalence of events was the highest. As compared to previous models, our approach has the advantage to consider the entire history of measurements during the patient follow-up, and to provide personalized prediction. We then assessed whether our model was able to identify at-risk patients in HEPATHER, a cohort that includes patients regardless of their fibrosis stage and reflects the current management of patients in clinical practice. In this real-life cohort, results showed that our model accurately stratifies the risk of LRC with good calibration.

Several works involving paired liver biopsies have demonstrated fibrosis regression after SVR in chronic hepatitis C [25, 26, 27, 28, 29], and even cirrhosis regression to lower fibrosis stages in 33% to 66% of cases [30]. Huang *et al.* recently reported that post-SVR fibrosis stage performed better than pre-treatment status for the long-term prediction of HCC [29]. In parallel and notwithstanding the dramatic decrease in FIB4 results after SVR, it seems that this simple blood test continues to correlate well with histological fibrosis stages after SVR [25, 26]. Taken together, these observations suggest that FIB4 may have prognostic value after SVR. Our results are consistent with this and, even more, we found that the FIB4 trajectory after SVR is associated with the residual risk of LRCs, as demonstrated by the rejection of the conditional independence assumption for the joint model with one latent class (**Supplementary Table S3**). Therefore, monitoring the dynamics of FIB4 is of interest after SVR in chronic hepatitis C.

Recent studies have evaluated the prognostic significance of FIB4 evolution after SVR [5, 6, 31]. In the Kanwal *et al.* work [6], the HCC risk in cirrhotic patients was higher in the subgroup with persistently high FIB4/APRI results at 12 months of follow-up, and lower in the subgroup with decreased FIB4/APRI results. The Ioannou *et al.* study reported similar findings [5]: a drop in FIB4 results from ≥ 3.25 pre-SVR to < 3.25 post-SVR was associated with a 50% reduced HCC risk in cirrhotic patients, but the absolute annual risk remained above 2% per year. These two works were however limited by the binary categorization of FIB4 results. That approach enabled the stratification of patients into four different subsets but not an evaluation of the prognostic significance of the magnitude of FIB4 evolution after SVR. By considering FIB4 as a continuous parameter and its trajectory over time, our JLCMs circumvent these limitations and furthermore provide the

additional advantage of individualized risk prediction.

A recent modelling study has found that the incremental cost-effectiveness ratio for biannual surveillance would exceed the traditional cost-effectiveness threshold if HCC incidence dropped to a value $<1.32\%$ per year [32]. Authors concluded that, due to the low incidence of HCC following SVR, identifying high-risk groups who are more likely to benefit from surveillance would be of high interest for tailoring guidelines. It has also been recently found that HCC screening by MRI could be more cost-effective than screening by ultrasonography in patients with an annual HCC risk $>3\%$ [33], suggesting that patient management modalities could be tailored to the individual risk level. In the present study, we provide physicians with a practical tool as a form of a web application to help them calculating the risk of LRC after SVR with very simple parameters. Such tool able to finely stratify the individual risk will be of importance for future works aimed at better refining the management of patients after SVR.

Our modelling found a subgroup with poor prognosis, who made up a very small subset of our population. Paradoxically, this subgroup experienced a high rate of LRCs despite a dramatic and rapid decrease in FIB4 (**Supplementary Figure S1**). This FIB4 decrease was explained more so by improvements in serum transaminases linked to the clearance of liver inflammation following SVR. In contrast, platelets remained low (below 150 G/L in many cases) and reflected the severity of the underlying liver disease and therefore the high LRC risk (**Supplementary Figure 2**). Further work is required to evaluate whether reinforced surveillance and management of these patients could help to improve their prognosis.

Several teams [5, 6, 31] have previously developed models for the prediction of LRCs after SVR, and interestingly, all of them identified FIB4 or its components as prognostic markers. However, those models only considered baseline parameters. The results published here demonstrate that our JLCM approach improves the accuracy of LRC risk predictions as more information is collected (**Table 3**). Indeed, calibration of our model for the 3-year risk prediction was better after accumulating the data of V0, V1 and V2 visits (one-year prediction at V2) compared with considering only baseline data (three-year prediction at V0). Such better individual risk adjustment with visit accumulation was even more pronounced in patients with poorer prognosis (predicted risk $\geq 1.5\%$) [34]. Some may argue that classical models developed using only baseline parameters can be easily repeated to reassess the risk during the patient follow-up, but we found this approach could be inappropriate. Indeed, with such « baseline » models, patients will be assigned to a low risk if FIB4 decreases during follow-up whereas, as discussed above, patients may remain at high risk of LRCs despite an important decrease in FIB4 and this because of a persistent severe underlying liver disease. Our JLCM approach has the strong advantage of adequately maintaining these patients in the high-risk category, as it considers both their initial status at baseline and the evolution of their parameters during follow-up.

In conclusion, the kinetics of FIB4 is associated with the residual risk of liver-related complications after sustained viral response in chronic hepatitis C. We developed a model considering FIB4 trajectories that provide very good accuracy for the individual prediction

of liver-related complications after SVR. Such model will improve personalized medicine in chronic hepatitis C and may serve to better identify patients for whom surveillance needs to be reinforced.

Acknowledgements

ANRS CO12 CirVir group (derivation cohort)*

Pierre Nahon¹, Tarik Asselah², Dominique Guyader³, Stanislas Pol⁴, H el ene Fontaine⁴, Georges-Philippe Pageaux⁵, Victor De L edinghen⁶, Denis Ouzan⁷, Fabien Zoulim⁸, Dominique Roulot⁹, Albert Tran¹⁰, Jean-Pierre Bronowicki¹¹, Thomas Decaens¹², Vincent Leroy¹², Ghassan Riachi¹³, Paul Cal es¹⁴, Jean-Marie P eron¹⁵, Laurent Alric¹⁶, Marc Bourli ere¹⁷, Philippe Mathurin¹⁸, Sebastien Dharancy¹⁸, Jean-Fr ed eric Blanc¹⁹, Armand Abergel²⁰, Olivier Chazouill eres²¹, Ariane Mallat²², Jean-Didier Grang e²³, Pierre Attali²⁴, Louis d'Alteroche²⁵, Claire Wartelle²⁶, Th ong Dao²⁷, Dominique Thabut²⁸, Christophe Pilette²⁹, Christine Silvain³⁰, Christos Christidis³¹, Eric Nguyen-Khac³², Brigitte Bernard-Chabert³³, David Zucman³⁴, Vincent Di Martino³⁵.

¹AP-HP, H opital Jean Verdier, Service d'H epatoLOGIE, Bondy, Universit e Paris 13, Bobigny et INSERM U1162, Universit e Paris 5, Paris; ²AP-HP, H opital Beaujon, Service d'H epatoLOGIE, Clichy; ³CHU Pontchaillou, Service d'H epatoLOGIE, Rennes; ⁴AP-HP, H opital Cochin, D epartement d'H epatoLOGIE et INSERM UMS20 et U1223, Institut Pasteur, Universit e Paris Descartes, Paris; ⁵H opital Saint Eloi, Service d'H epatoLOGIE, Montpellier; ⁶H opital Haut-L ev eque, Service d'H epatoLOGIE, Bordeaux; ⁷Institut Arnaud Tzanck, Service d'H epatoLOGIE, St Laurent du Var; ⁸H opital H otel Dieu, Service d'H epatoLOGIE, Lyon; ⁹AP-HP, H opital Avicenne, Service d'H epatoLOGIE, Bobigny; ¹⁰CHU de Nice, Service d'H epatoLOGIE, et INSERM U1065, Universit e de Nice-Sophia-Antipolis, Nice; ¹¹H opital Brabois, Service d'H epatoLOGIE, Vandoeuvre-les-Nancy; ¹²H opital Michallon, Service d'H epatoLOGIE, Grenoble; ¹³H opital Charles-Nicolle, Service d'H epatoLOGIE, Rouen; ¹⁴CHU d'Angers, Service d'H epatoLOGIE, Angers; ¹⁵H opital Purpan, Service d'H epatoLOGIE, Toulouse; ¹⁶CHU Toulouse, Service de M edecine Interne-P ole Digestif UMR 152, Toulouse; ¹⁷H opital Saint Joseph, Service d'H epatoLOGIE, Marseille; ¹⁸H opital Claude Huriez, Service d'H epatoLOGIE, Lille; ¹⁹H opital St Andr e, Service d'H epatoLOGIE, Bordeaux; ²⁰H opital H otel Dieu, Service d'H epatoLOGIE, Clermont-Ferrand; ²¹AP-HP, H opital Saint-Antoine, Service d'H epatoLOGIE, Paris; ²²AP-HP, H opital Henri Mondor, Service d'H epatoLOGIE, Cr eteil; ²³AP-HP, H opital Tenon, Service d'H epatoLOGIE, Paris; ²⁴AP-HP, H opital Paul Brousse, Service d'H epatoLOGIE, Villejuif; ²⁵H opital Trousseau, Unit e d'H epatoLOGIE, CHRU de Tours; ²⁶H opital d'Aix-En-Provence, Service d'H epatoLOGIE, Aix-En-Provence; ²⁷H opital de la C ote de Nacre, Service d'H epatoLOGIE, Caen; ²⁸Paris Sorbonne Universit e, AP-HP, Groupe Hospitalier de La Piti e-Salp etri ere, Service d'H epatoLOGIE, Paris; ²⁹CHU Le Mans, Service d'H epatoLOGIE, Le Mans; ³⁰CHU de Poitiers, Service d'H epatoLOGIE, Poitiers; ³¹Institut Mutualiste Montsouris, Service d'H epatoLOGIE, Paris; ³²H opital Amiens Nord, Service d'H epatoLOGIE, Amiens; ³³H opital Robert Debr e, Service d'H epatoLOGIE, Reims; ³⁴H opital Foch, Service de M edecine Interne, Suresnes; ³⁵H opital Jean Minjoz, Service d'H epatoLOGIE, Besan on. France.

Scientific Committee P. Nahon (principal investigator), R. Layese and E Audureau (data management), P. Bedossa, M. Bonjour, V. Bourcier, S. Dharancy, I. Durand-Zaleski, H. Fontaine, D. Guyader, A. Laurent, V. Leroy, P. Marche, D. Salmon, V. Thibault, V. Vilgrain, J. Zucman-Rossi, C. Cagnot (ANRS), V. Petrov-Sanchez (ANRS).

Clinical centers (Department/participating physicians) CHU Jean Verdier, Bondy (P. Nahon, V. Bourcier); CHU Cochin, Paris (S. Pol, H. Fontaine); CHU Pitié-Salpêtrière, Paris (D. Thabut); CHU Saint-Antoine, Paris (L. Serfaty); CHU Avicenne, Bobigny (D. Roulot); CHU Beaujon, Clichy (P. Marcellin); CHU Henri Mondor (A. Mallat); CHU Paul Brousse (P. Attali); CHU Tenon, Paris (J.D. Grangé); CHRU Hôpital Nord, Amiens (D. Capron); CHU Angers (P. Calès); Hôpital Saint-Joseph, Marseille (M. Bourlière); CHU Brabois, Nancy (J.P. Bronowicki); Hôpital Archet, Nice (A. Tran); Institut Mutualiste Montsouris, Paris (F. Mal, C. Christidis); CHU Poitiers (C. Silvain); CHU Pontchaillou, Rennes (D. Guyader); CH Pays d'Aix, Aix-en-Provence (C. Wartelle); CHU Jean Minjot, Besancon (V. Di Martino); CHU Bordeaux - Hôpital Haut-Leveque, Pessac (V. de Ledinghen); CHU Bordeaux - Hôpital Saint-André, Bordeaux (JF. Blanc); CHU Hôtel Dieu, Lyon (C. Trepo, F. Zoulim); CHU Clermont-Ferrand (A. Abergel); Hôpital Foch, Suresnes (S. Hillaire); CHU Caen (T. Dao); CHU Lille (P. Mathurin); CH Le Mans (C. Pilette); CHU Michallon, Grenoble (JP. Zarski); CHU St Eloi, Montpellier (D. Larrey); CHU Reims (B. Bernard-Chabert); CHU Rouen (O. Gorla, G. Riachi); Institut Arnaud Tzanck, St Laurent-du-Var (D. Ouzan); CHU Purpan, Toulouse (JM. Péron, L. Alric); CHU Tours (Y. Bacq).

* This study was sponsored by the ANRS (France REcherche Nord & sud SIDA-HIV Hépatites: FRENH). This work is dedicated to the memory of Professor Jean-Claude Trinchet.

ANRS CO22 HEPATHER group (validation cohort)**

ANRS-AFEF HPATHER study group Delphine Bonnet, Virginie Payssan-Sicart, Chloe Pomes (CHU Purpan, Toulouse, France), François Bailly, Marjolaine Beaudoin, Dominique Giboz, Kerstin Hartig-Lavie, Marianne Maynard (Hospices Civils de Lyon, Lyon, France), Eric Billaud, David Bouteille, Morane Cavellec, Marjorie Cheraud-Carpentier (Hôpital Hôtel-Dieu, Nantes, France), Isabelle Hubert, Jaouad Benhida, Adrien Lannes, Françoise Lunel, Frédéric Oberti (CHU Angers, Angers, France), Nathalie Boyer, Nathalie Giully, Corinne Castelnau, Giovanna Scoazec (Hôpital Beaujon, Clichy, France), Aziza Chibah, Sylvie Keser, Karim Bonardi, Anaïs Vallet-Pichard, Philippe Sogni (Hôpital Cochin, Paris, France), Juliette Foucher, Jean-Baptiste Hiriart, Amy Wilson, Sarah Shili, Faiza Chermak (Hôpital Haut-Lévêque, Pessac, Bordeaux, France), Christelle Ansaldi, Nissérine Ben Amara, Laëtitia Chouquet, Emilie De Luca, Valérie Oules (Hôpital Saint Joseph, Marseille, France), Rodolphe Anty, Eve Gelsi, Régine Truchi (CHU de Nice, Nice, France), Elena Luckina, Nadia Messaoudi, Joseph Moussali (Hôpital de la Pitié Salpêtrière, Paris, France), Barbara De Dieuleveult, Damien Labarriere, Pascal Poter, Si Nafa Si Ahmed (CHR La Source, Orléans, France), Véronique Grando-Lemaire, Pierre Nahon, Valérie Bourcier, Séverine Brulé, Thomas Stallberger (Hôpital Jean Verdier, Bondy, France), Caroline Jezequel, Audrey Brener, Anne Laligant, Aline Rabot, Isabelle Renard (CHU

Rennes, Rennes, France), Thomas F Baumert, Michel Doffoel, Catherine Mutter, Pauline Simo-Noumbissie, Esma Razi (Hôpitaux Universitaires de Strasbourg, Strasbourg, France), Hélène Barraud, Mouni Bensenane, Abdelbasset Nani, Sarah Hassani-Nani, Marie-Albertine Bernard (CHU de Nancy, Nancy, France), Georges-Philippe Pageaux, Michael Bismuth, Ludovic Caillo, Stéphanie Faure, Marie Pierre Ripault (Hôpital Saint Eloi, Montpellier, France), Christophe Bureau, Jean Marie Peron, Marie Angèle Robic, Léa Tarallo (CHU Purpan, Toulouse, France), Marine Faure, Bruno Froissart, Marie-Noëlle Hilleret, Jean-Pierre Zarski (CHU de Grenoble, Grenoble, France), Odile Gorla, Victorien Grard, Hélène Montialoux (CHU Charles Nicolle, Rouen, France), Muriel François, Christian Ouedraogo, Christelle Pauleau, Anne Varault (Hôpital Henri Mondor, Créteil, France), Tony Andreani, Bénédicte Angoulevant, Azeline Chevance, Lawrence Serfaty (Hôpital Saint-Antoine, Paris, France), Teresa Antonini, Audrey Coilly, Jean-Charles Duclos Vallée, Mariagrazia Tateo (Hôpital Paul Brousse, Villejuif, France), Corinne Bonny, Chanteranne Brigitte, Géraldine Lamblin, Léon Muti (Hôpital Estaing, Clermont-Ferrand, France), Abdenour Babouri, Virginie Filipe (Centre Hospitalier Régional, Metz, France), Camille Barrault, Laurent Costes, Hervé Hagège, Soraya Merbah (Centre Hospitalier Intercommunal, Créteil, France), Paul Carrier, Maryline Debette-Gratien, Jérémie Jacques (CHU Limoges, Limoges, France), Guillaume Lassailly, Florent Artu, Valérie Canva, Sébastien Dharancy, Alexandre Louvet (CHRU Claude Huriez, Lille, France), Marianne Latournerie, Marc Bardou, Thomas Mouillot (Dijon University Hospital, Dijon, France), Yannick Bacq, Didier Barbereau, Charlotte Nicolas (CHU Trousseau, 37044 Tours, France), Caroline Chevalier, Isabelle Archambeaud, Sarah Habes (CHU de Nantes, Nantes, France), Nisserine Ben Amara, Danièle Botta-Fridlund, (CHU Timone, Marseille, France), Eric Saillard, Marie-Josée Lafrance, (CHU de Pointe-à-Pitre, Pointe-à-Pitre, Guadeloupe).

Scientific committee - voting members Marc Bourlière (Hôpital St Joseph, Marseille), Patrice Cacoub (Hôpital Pitié salpêtrière, Paris, France), Fabrice Carrat (Scientific Coordinator, Hôpital Saint-Antoine, Paris, France), Patrizia Carrieri (INSERM U912, Marseille, France), Elisabeth Delarocque-Astagneau (Inserm UMR1181, Paris), Victor De Ledinghen (Hôpital Haut-Lévêque, Pessac, Bordeaux, France), Céline Dorival (UPMC & INSERM U1136, Paris, France), Jean Dubuisson (Inserm U1019, Lille, France), Hélène Fontaine (Hôpital Cochin, Paris, France), Chantal Housset (Inserm UMR-S938 1 IFR65, Paris), Dominique Larrey (Hôpital Saint Eloi, Montpellier, France), Patrick Marcellin (Hôpital Beaujon, Clichy, France), Philippe Mathurin (CHRU Claude Huriez, Lille, France), Pierre Nahon (Hôpital Jean Verdier, Bondy, France), Georges-Philippe Pageaux (Hôpital Saint Eloi, Montpellier, France), Jean-Michel Pawlotsky (Hôpital Henri Mondor, Créteil, France), Ventzislava Petrov-Sanchez (ANRS, Paris, France), Stanislas Pol (Principal Investigator, Hôpital Cochin, Paris, France), Sophie Vaux (Agence Nationale de Santé Publique, Saint Maurice, France), Linda Wittkop (ISPED-INSERM U897, Bordeaux, France), Yazdan Yazdanpanah (Hôpital Bichat Claude Bernard, Paris, France), Jean-Pierre Zarski (CHU de Grenoble, Grenoble, France), Fabien Zoulim (Hospices Civils de Lyon, Lyon, France), Jessica Zucman-Rossi (INSERM U6741162, Paris).

Scientific committee - non-voting members Marianne L'hennaff (ARCAT-TRT-5-CHV, France), Michèle Sizorn (SOS hépatites, France); one representative of INSERM-

ANRS Pharmacovigilance team, Paris, France (Imane Amri, Alpha Diallo), Mélanie Simony, Carole Cagnot (INSERM-ANRS, Paris, France), one member of INSERM Transfert, Paris, France (Alice Bousselet, Mireille Caralp, Jean-Marc Lacombe), and one representative of each pharmaceutical company (MSD, Janssen, Gilead, AbbVie, Bristol-Myers Squibb, Roche).

** Financial support provided by ANRS (France REcherche Nord & sud SIDA-HIV Hépatites: FRENESH), ANR (Agence Nationale de la Recherche), DGS (Direction Générale de la Santé), MSD, Janssen, Gilead, AbbVie, Bristol-Myers Squibb, and Roche. **Sponsor:** Imane Amri, Alpha Diallo, Carole Cagnot, Mélanie Simony (INSERM-ANRS, Paris, France), Ventzi Petrov-Sanchez (coordinator). **Methodology and Coordinating Centre:** Loubna Ayour, Jaouad Benhida, Fabrice Carrat (coordinator), Frederic Chau, Céline Dorival, Audrey Gilibert, Isabelle Goderel, Victorien Gard, Warda Hadi, Georges Haour, Godwin Mawuvi, Léa Mba Mintsá, Grégory Pannetier, François Pinot, Muriel Sudres, François Téloulé (Sorbonne Université & INSERM U1136, Paris, France).

References

- [1] Spearman W, Dusheiko G, Hellard M, Sonderup M. Hepatitis C. *The Lancet* 2019; 394(10207): 1451-1466.
- [2] Carrat F, Fontaine H, Dorival C, et al. Clinical outcomes in patients with chronic hepatitis C after direct-acting antiviral treatment: a prospective cohort study. *The Lancet* 2019; 393(10179): 1453-1464.
- [3] Backus L, Belperio P, Shahoumian T, Mole L. Impact of sustained virologic response with direct-acting antiviral treatment on mortality in patients with advanced liver disease. *Hepatology* 2019; 69(2): 487-497.
- [4] Sahakyan Y, Lee-Kim V, Bremner KE, Bielecki JM, Krahn MD. Impact of direct-acting antiviral regimens on mortality and morbidity outcomes in patients with chronic hepatitis c: Systematic review and meta-analysis. *Journal of Viral Hepatitis* 2021; 28(5): 739-754.
- [5] Ioannou G, Beste L, Green P, et al. Increased risk for hepatocellular carcinoma persists up to 10 years after HCV eradication in patients with baseline cirrhosis or high FIB-4 scores. *Gastroenterology* 2019; 157(5): 1264-1278.
- [6] Kanwal F, Kramer JR, Asch SM, Cao Y, Li L, El-Serag HB. Long-term risk of hepatocellular carcinoma in HCV patients treated with direct acting antiviral agents. *Hepatology* 2020; 71(1): 44-55.
- [7] Bhagat N, Verma N, Singh V. HCC prediction post SVR: Many tools yet limited generalizability!. *Journal of Hepatology* 2022.
- [8] Pons M, Rodríguez-Tajes S, Esteban JI, et al. Non-invasive prediction of liver-related events in patients with HCV-associated compensated advanced chronic liver disease after oral antivirals. *Journal of Hepatology* 2020; 72(3): 472-480.

-
- [9] Alonso Lopez S, Manzano ML, Gea F, et al. A model based on noninvasive markers predicts very low hepatocellular carcinoma risk after viral response in hepatitis C virus–advanced fibrosis. *Hepatology* 2020; 72(6): 1924–1934.
- [10] Trinchet JC, Bourcier V, Chaffaut C, et al. Complications and competing risks of death in compensated viral cirrhosis (ANRS CO12 CirVir prospective cohort). *Hepatology* 2015; 62(3): 737–750.
- [11] Wai CT, Greenson J, Fontana R, et al. A simple noninvasive index can predict both significant fibrosis and cirrhosis in patients with chronic hepatitis C. *Hepatology* 2003; 38(2): 518–526.
- [12] Sterling R, 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(6): 1317–1325.
- [13] Blanche P, Dartigues JF, Jacqmin-Gadda H. Estimating and comparing time-dependent areas under receiver operating characteristic curves for censored event times with competing risks. *Statistics in Medicine* 2013; 32(30): 5381–5397.
- [14] Proust-Lima C, Taylor J. Development and validation of a dynamic prognostic tool for prostate cancer recurrence using repeated measures of post-treatment PSA: a joint modeling approach. *Biostatistics* 2009; 10(3): 535–549.
- [15] Graf E, Schmoor C, Sauerbrei W, Schumacher M. Assessment and comparison of prognostic classification schemes for survival data. *Statistics in medicine* 1999; 18(17-18): 2529–2545.
- [16] Proust-Lima C, Sène M, Taylor JM, Jacqmin-Gadda H. Joint latent class models for longitudinal and time-to-event data: a review. *Statistical Methods in Medical research* 2014; 23(1): 74–90.
- [17] Ghany M, Morgan T, AASLD-IDSAs. Hepatitis C guidance 2019 update: American Association for the Study of Liver Diseases–Infectious Diseases Society of America recommendations for testing, managing, and treating hepatitis C virus infection. *Hepatology* 2020; 71(2): 686–721.
- [18] Pawlotsky JM, Negro F, Aghemo A, et al. EASL recommendations on treatment of hepatitis C: final update of the series. *Journal of Hepatology* 2020; 73(5): 1170–1218.
- [19] Zarski JP, Sturm N, Guechot J, et al. Comparison of nine blood tests and transient elastography for liver fibrosis in chronic hepatitis C: the ANRS HCEP-23 study. *Journal of Hepatology* 2012; 56(1): 55–62.
- [20] Degos F, Perez P, Roche B, et al. Diagnostic accuracy of FibroScan and comparison to liver fibrosis biomarkers in chronic viral hepatitis: a multicenter prospective study (the FIBROSTIC study). *Journal of Hepatology* 2010; 53(6): 1013–1021.

-
- [21] Bachofner JA, Valli PV, Kröger A, et al. Direct antiviral agent treatment of chronic hepatitis C results in rapid regression of transient elastography and fibrosis markers fibrosis-4 score and aspartate aminotransferase-platelet ratio index. *Liver International* 2017; 37(3): 369–376.
- [22] Singh S, Facciorusso A, Loomba R, Falck-Ytter YT. Magnitude and kinetics of decrease in liver stiffness after antiviral therapy in patients with chronic hepatitis C: a systematic review and meta-analysis. *Clinical Gastroenterology and Hepatology* 2018; 16(1): 27–38.
- [23] Azzi J, Dorival C, Cagnot C, et al. Prediction of hepatocellular carcinoma in Hepatitis C patients with advanced fibrosis after sustained virologic response. *Clinics and Research in Hepatology and Gastroenterology* 2022: 101923.
- [24] Poynard T, Lacombe JM, Deckmyn O, et al. External validation of LCR1-LCR2, a multivariable HCC risk calculator, in patients with chronic HCV. *JHEP Reports* 2021; 3(4): 100298.
- [25] Huang R, Rao H, Yang M, et al. Noninvasive measurements predict liver fibrosis well in hepatitis C virus patients after direct-acting antiviral therapy. *Digestive Diseases and Sciences* 2020; 65(5): 1491–1500.
- [26] Tachi Y, Hirai T, Toyoda H, et al. Predictive ability of laboratory indices for liver fibrosis in patients with chronic hepatitis C after the eradication of hepatitis C virus. *PLoS One* 2015; 10(7).
- [27] Wang X, Gao F, Yuan G, et al. Ten-year follow-up analysis of chronic hepatitis C patients after getting sustained virological response to pegylated interferon- α and ribavirin therapy. *Journal of Viral Hepatitis* 2016; 23(12): 971–976.
- [28] Shiratori Y, Imazeki F, Moriyama M, et al. Histologic improvement of fibrosis in patients with hepatitis C who have sustained response to interferon therapy. *Annals of Internal Medicine* 2000; 132(7): 517–524.
- [29] Huang CF, Yeh ML, Huang CI, et al. Post-treatment fibrotic modifications overwhelm pre-treatment liver fibrosis in predicting HCC in CHC patients with curative antivirals. *Hepatology International* 2018; 12(6): 544–551.
- [30] D’Ambrosio R, Degasperi E, Aghemo A, et al. Serological tests do not predict residual fibrosis in hepatitis C cirrhotics with a sustained virological response to interferon. *PLoS One* 2016; 11(6).
- [31] Toyoda H, Tada T, Yasuda S, Mizuno K, Ito T, Kumada T. Dynamic evaluation of liver fibrosis to assess the risk of hepatocellular carcinoma in patients with chronic hepatitis C who achieved sustained virologic response. *Clinical Infectious Diseases* 2020; 70(6): 1208–1214.
- [32] Zangneh HF, Wong W, Sander B, et al. Cost effectiveness of hepatocellular carcinoma surveillance after a sustained virologic response to therapy in patients with hepatitis C

-
- virus infection and advanced fibrosis. *Clinical Gastroenterology and Hepatology* 2019; 17(9): 1840–1849.
- [33] Nahon P, Najean M, Layese R, et al. Early hepatocellular carcinoma detection using magnetic resonance imaging is cost-effective in high-risk patients with cirrhosis. *JHEP Reports* 2022; 4(1): 100390.
- [34] Blanche P, Proust-Lima C, Loubère L, Berr C, Dartigues JF, Jacqmin-Gadda H. Quantifying and comparing dynamic predictive accuracy of joint models for longitudinal marker and time-to-event in presence of censoring and competing risks. *Biometrics* 2015; 71(1): 102–113.

Table legend

Table 1: Baseline characteristics of the derivation and validation study sets

Table 2: Cumulative observed incidence of LRC at 1, 2 and 3 years, as a function of patient groups defined by the predicted risk of LRC by the study model

Predicted risk of LRC by the study model $<1.5\%$ per year supposes the observed incidence of LRC at one year will be $<1.5\%$, $<3.0\%$ at two years, and $<4.5\%$ at 3 years. Similarly, predicted risk between $1.5-3.0\%$ per year supposes the observed incidence of LRC will be within $1.5-3.0\%$ at one year, $3.0-6.0\%$ at two years, and $4.5-9.0\%$ at three years. Finally, predicted risk $\geq 3.0\%$ per year supposes the observed incidence of LRC will be $\geq 3.0\%$ at one year, $\geq 6.0\%$ at two years, and $\geq 9.0\%$ at three years. Correct predictions are noted in bold.

Table 3: Time-dependent Brier Score for the prediction of liver-related complications by the study model at different visits and different time horizons

Predictions of LRC risk at 2, 3 and 4-year horizons can be calculated in different ways. Brier Scores in blue indicate a 2-year horizon (« Baseline visit V0 + 2 years » or « One-year visit V1 + 1 year »), in green a 3-year horizon (« Baseline visit V0 + 3 years », « One-year visit V1 + 2 years » or « Two-year visit V2 + 1 year »), and in red a 4-year horizon (« One-year visit V1 + 3 years » or « Two-year visit V2 + 2 years »).

Table 1: Baseline characteristics of the derivation and validation study sets

	Derivation set (n=695)	Validation set (n=7,064)
Baseline:		
Age (years)	57.4 (51.0-64.6)	58.2 (52.3-66.3)
Male sex (%)	456 (65.4)	3790 (53.7)
BMI (kg/m ²)	25.7 (22.9-29.1)	25.0 (22.4-28.0)
Diabetes (%)	123 (17.7)	893 (12.6)
Arterial hypertension (%)	197 (28.3)	2281 (32.3)
Dyslipidemia (%)	48 (6.9)	100 (1.4)
Prothrombin time (%)	90 (81-99)	95 (87-100)
Bilirubin (μ mol/L)	10.0 (6.8-14.0)	10.0 (7.0-14.1)
Albumin (g/L)	43.0 (40.0-46.0)	43.0 (40.0-45.4)
Platelets (G/L)	154 (107-200)	202 (160-245)
AST (IU/L)	30 (24-39)	24 (19-29)
ALT (IU/L)	28 (21-40)	20 (15-28)
GGT (IU/L)	44 (27-76)	25 (17-42)
APRI	0.52 (0.32-0.86)	0.29 (0.21-0.42)
FIB4	2.24 (1.47-3.54)	1.55 (1.11-2.18)
Follow-up:		
Follow-up time (years)	3.8 (1.6-7.5)	3.6 (2.5-4.9)
Number of visits (n)	6 (4-11)	5 (3-6)
HCC (%)	25 (3.6)	204 (2.9)
Cirrhosis decompensation (%)	25 (3.6)	69 (1.0)

Continuous variables are expressed as median with 1st and 3rd quartiles into bracket, and categorical data as number of patients with frequency (%) into bracket.

BMI: Body Mass Index; AST: ASpartate aminoTransferase; ALT: ALanine aminoTransferase; GGT: Gamma-GT; HCC: HepatoCellular Carcinoma.

Table 2: Cumulative observed incidence of LRC at 1, 2 and 3 years, as a function of patient groups defined by the predicted risk of LRC by the study model

Predicted risk of LRC by the study model <1.5% per year supposes the observed incidence of LRC at one year will be <1.5%, <3.0% at two years, and <4.5% at 3 years. Similarly, predicted risk between 1.5-3.0% per year supposes the observed incidence of LRC will be within 1.5-3.0% at one year, 3.0-6.0% at two years, and 4.5-9.0% at three years. Finally, predicted risk \geq 3.0% per year supposes the observed incidence of LRC will be \geq 3.0% at one year, \geq 6.0% at two years, and \geq 9.0% at three years. Correct predictions are noted in bold.

Visit	At 1 year		At 2 years		At 3 years	
	Predicted	Observed	Predicted	Observed	Predicted	Observed
Baseline V0	<1.5%	0.4 (0.2-0.6)	<3.0%	1.0 (0.7-1.3)	<4.5%	1.7 (1.3-2.1)
	1.5-3.0%	1.5 (0.8-2.3)	3.0-6.0%	3.3 (2.1-4.4)	4.5-9.0%	5.1 (3.7-6.5)
	\geq 3.0%	5.2 (2.6-7.7)	\geq 6.0%	11.9 (8.1-15.6)	\geq 9%	16.5 (12.0-20.8)
One-year V1	<1.5%	0.5 (0.3-0.7)	<3.0%	1.4 (1.0-1.7)	<4.5%	2.5 (2.0-3.1)
	1.5-3.0%	2.2 (1.2-3.2)	3.0-6.0%	4.1 (2.7-5.5)	4.5-9.0%	5.3 (3.6-6.9)
	\geq 3.0%	8.1 (4.7-11.3)	\geq 6.0%	12.2 (8.1-16.2)	\geq 9%	19.9 (14.5-25.0)
Two-year V2	<1.5%	1.0 (0.7-1.4)	<3.0%	2.0 (1.4-2.5)	<4.5%	2.8 (2.1-3.4)
	1.5-3.0%	2.1 (1.0-3.3)	3.0-6.0%	2.9 (1.5-4.3)	4.5-9.0%	4.3 (2.4-6.2)
	\geq 3.0%	5.0 (2.1-7.8)	\geq 6.0%	12.5 (7.8-17.1)	\geq 9%	19.6 (13.3-25.3)

Table 3: Time-dependent Brier Score for the prediction of liver-related complications by the study model at different visits and different time horizons

Predictions of LRC risk at 2, 3 and 4-year horizons can be calculated in different ways. Brier Scores in blue indicate a 2-year horizon (« Baseline visit V0 + 2 years » or « One-year visit V1 + 1 year »), in green a 3-year horizon (« Baseline visit V0 + 3 years », « One-year visit V1 + 2 years » or « Two-year visit V2 + 1 year »), and in red a 4-year horizon (« One-year visit V1 + 3 years » or « Two-year visit V2 + 2 years »).

Visit	Predicted	Time-dependent Brier Score		
	risk per year	One-year horizon	Two-year horizon	Three-year horizon
Baseline V0	All	0.007 (0.005-0.009)	0.016 (0.013-0.019)	0.023 (0.020-0.026)
	<1.5%	0.004 (0.002-0.005)	0.009 (0.007-0.011)	0.014 (0.011-0.017)
	1.5-3.0%	0.015 (0.008-0.022)	0.030 (0.020-0.040)	0.044 (0.032-0.055)
	≥3.0%	0.048 (0.027-0.071)	0.100 (0.076-0.128)	0.131 (0.105-0.156)
One-year V1	All	0.011 (0.008-0.014)	0.020 (0.017-0.024)	0.030 (0.025-0.034)
	<1.5%	0.005 (0.003-0.008)	0.011 (0.008-0.015)	0.019 (0.015-0.023)
	1.5-3.0%	0.022 (0.012-0.036)	0.041 (0.026-0.057)	0.050 (0.034-0.067)
	≥3.0%	0.081 (0.049-0.121)	0.117 (0.083-0.154)	0.156 (0.121-0.188)
Two-year V2	All	0.013 (0.010-0.017)	0.023 (0.018-0.029)	0.029 (0.023-0.034)
	<1.5%	0.011 (0.008-0.016)	0.018 (0.013-0.023)	0.021 (0.016-0.027)
	1.5-3.0%	0.016 (0.005-0.027)	0.023 (0.012-0.036)	0.028 (0.016-0.041)
	≥3.0%	0.039 (0.019-0.068)	0.093 (0.063-0.127)	0.136 (0.106-0.171)

Figure legend

Figure 1: Flow-charts of the derivation and validation study sets

SVR: Sustained Viral Response; LRC: Liver-Related Complication; HCC: HepatoCellular Carcinoma

Figure 2: Survival free of liver-related complication in the derivation (black) and validation (red) sets

Figure 3: Time-dependent area under receiver operating characteristic curves (AUC(t)) for the prediction of liver-related complications in the derivation set (CirVir cohort)

PLT: platelets; AST: ASpartate aminoTransferase; ALT: ALanine aminoTransferase

Figure 4: Observed cumulative incidence of Liver-Related Complication (LRC) in the validation set (HEPATHER cohort), as a function of patient groups defined by the predicted risk estimated by the study model

The predicted risk of LRC estimated by the study model was calculated for each patient at different visits (panel a: V0 baseline visit; panel b: V1 one-year follow-up visit; panel c: V2 two-year follow-up visit). Patient were then pooled in three groups according to their predicted risk: $<1.5\%$ per year (green), $1.5\text{-}3.0\%$ (yellow), and $\geq 3.0\%$ (red). Observed cumulative incidence curves of LRC were finally drawn for each group. The predicted risk is computed over the 3-year horizon.

Figure 5: Shiny application (available at <https://hifih.shinyapps.io/lrcrisk/>) for the evaluation in clinical practice of the risk of liver-related complications following sustained viral response

The user inputs the data in the application (date of birth, date of visit, gender, AST, ALT, platelets, diabetes status) at each follow-up visit. The results of the FIB4 are plotted on the left of the figure, the cumulative incidence of liver-related complications is calculated and drawn on the right (solid line, with 95% confidence interval in dashed lines). The meters at the top indicate the individual risk of liver-related complications at 1, 3 and 5 years. The panels depict an example for a patient at different time points. **Panel a:** risk of liver related complications calculated from FIB4 measured at sustained viral response and 1 year after; **Panel b:** risk calculated from the evolution of FIB4 until year 11.

Figure 1: Flow-charts of the derivation and validation study sets
 SVR: Sustained Viral Response; LRC: Liver-Related Complication; HCC: HepatoCellular Carcinoma

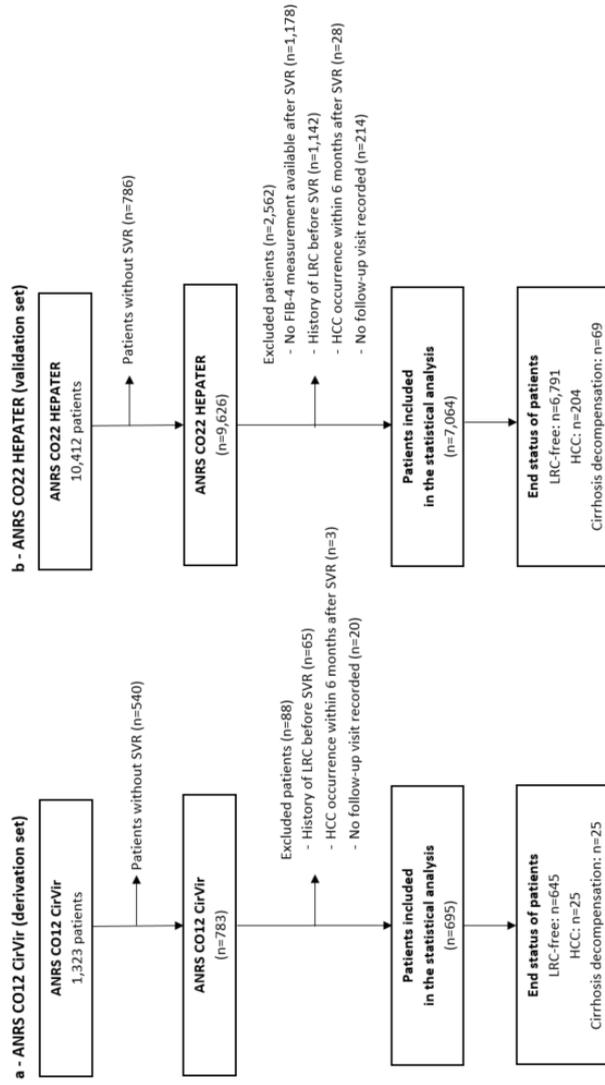


Figure 2: Survival free of liver-related complication in the derivation (black) and validation (red) sets

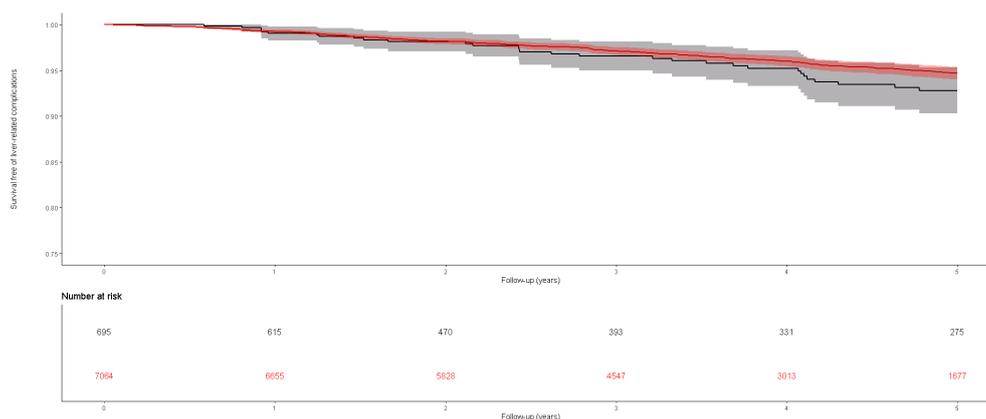


Figure 3: Time-dependent area under receiver operating characteristic curves (AUC(t)) for the prediction of liver-related complications in the derivation set (CirVir cohort)

PLT: platelets; AST: ASpartate aminoTransferase; ALT: ALanine aminoTransferase

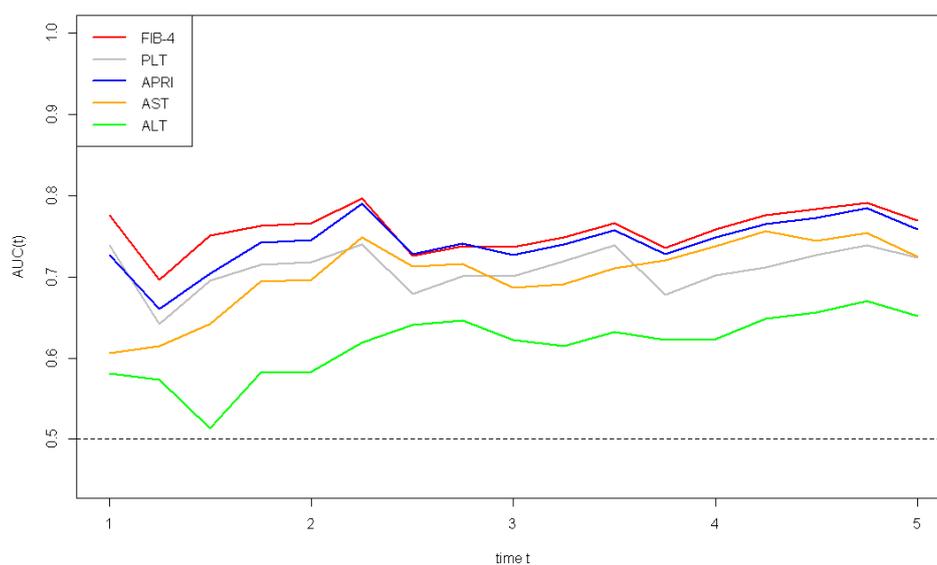
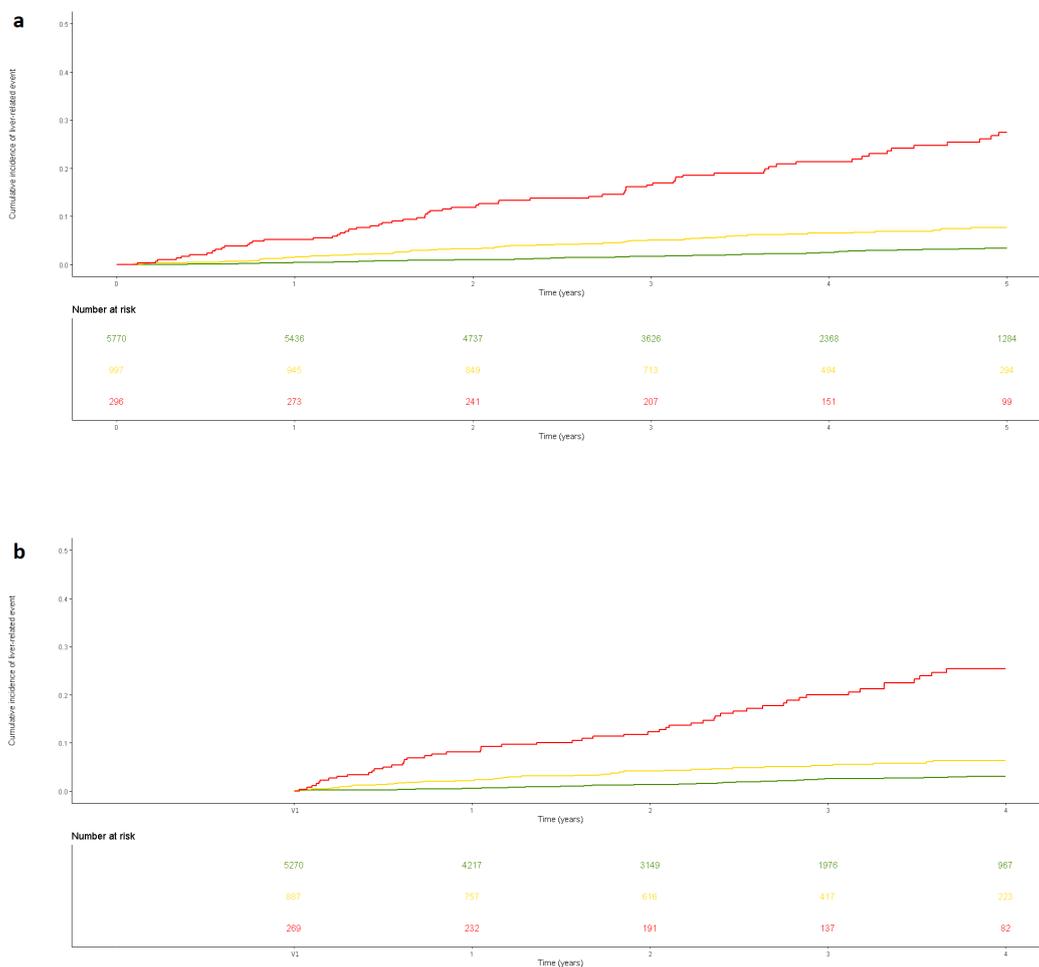


Figure 4: Observed cumulative incidence of Liver-Related Complication (LRC) in the validation set (HEPATHER cohort), as a function of patient groups defined by the predicted risk estimated by the study model

The predicted risk of LRC estimated by the study model was calculated for each patient at different visits (panel a: V0 baseline visit; panel b: V1 one-year follow-up visit; panel c: V2 two-year follow-up visit). Patient were then pooled in three groups according to their predicted risk: <1.5% per year (green), 1.5-3.0% (yellow), and $\geq 3.0\%$ (red). Observed cumulative incidence curves of LRC were finally drawn for each group. The predicted risk is computed over the 3-year horizon.



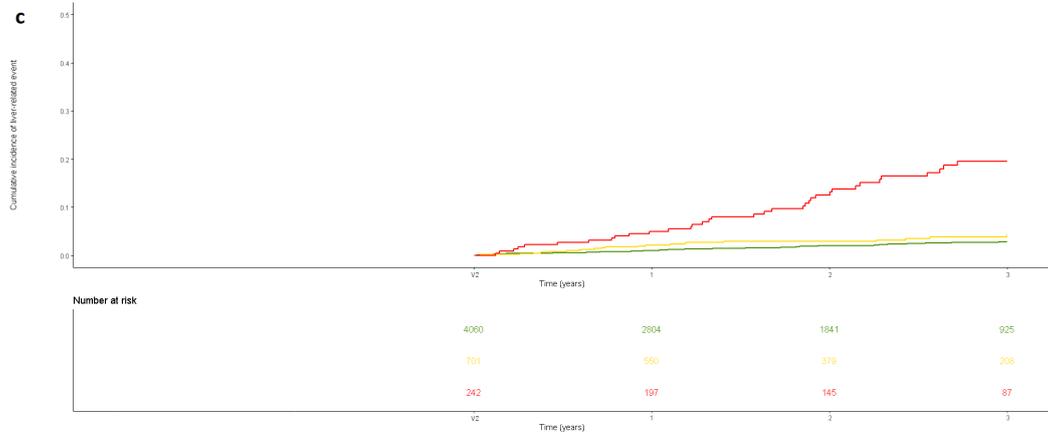
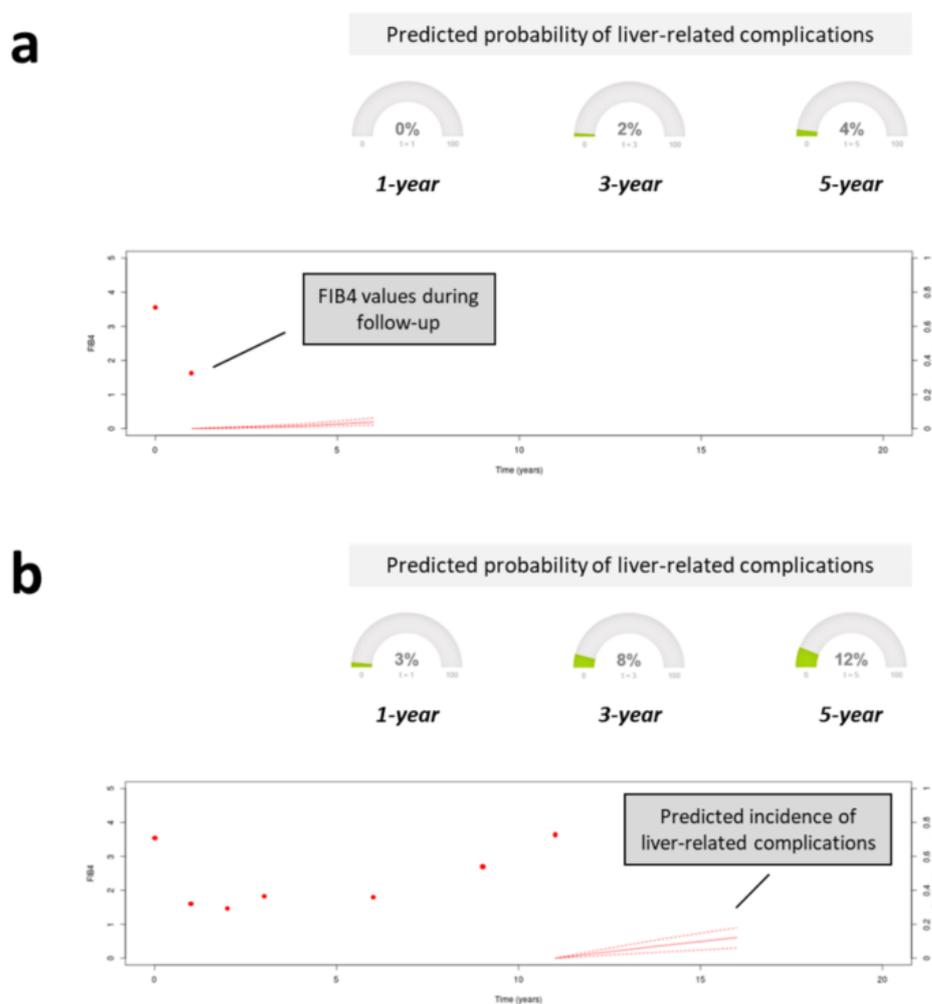


Figure 5: Shiny application (available at <https://hifih.shinyapps.io/lrcrisk/>) for the evaluation in clinical practice of the risk of liver-related complications following sustained viral response

The user inputs the data in the application (date of birth, date of visit, gender, AST, ALT, platelets, diabetes status) at each follow-up visit. The results of the FIB4 are plotted on the left of the figure, the cumulative incidence of liver-related complications is calculated and drawn on the right (solid line, with 95% confidence interval in dashed lines). The meters at the top indicate the individual risk of liver-related complications at 1, 3 and 5 years. The panels depict an example for a patient at different time points. **Panel a:** risk of liver related complications calculated from FIB4 measured at sustained viral response and 1 year after; **Panel b:** risk calculated from the evolution of FIB4 until year 11.



4.2 Documents complémentaires

Dynamic Personalized Prediction of the Individual Liver-Related Risk after Sustained Viral Response in HCV patients

Clémence Moreau¹, Marine Roux¹, Jérémie Riou^{2,3}, Pierre Nahon^{4,5},
Fabrice Carrat^{6,7}, Jérôme Boursier^{1,8}

¹Laboratoire HIFIH, UPRES EA3859, SFR 4208, Université d'Angers, Angers, France

²Centre Hospitalier Universitaire d'Angers, Délégation à la recherche clinique et à
l'innovation, Angers, France

³Inserm, UMR 1066, Micro et Nanomédecines translationnelles, Université d'Angers,
Angers, France

⁴AP-HP, Hôpital Jean Verdier, Unité d'Hépatologie, Bondy, France

⁵Inserm, UMR 1162, Génomique fonctionnelle des tumeurs solides, F-75000, Paris,
France

⁶AP-HP, Hôpital Saint-Antoine, Unité de Santé Publique, Paris, France

⁷Inserm, UMR-S 1136, Institut Pierre-Louis d'Epidémiologie et de Santé Publique,
Paris, France

⁸Centre Hospitalier Universitaire d'Angers, Service d'Hépto-Gastroentérologie et
Oncologie Digestive, Angers, France

SUPPLEMENTARY MATERIAL

Supplementary methods

Data collection during the follow-up in the CirVir and the HEP-ATHER cohorts

CirVir cohort

Patients were seen by physicians every six months as per cirrhosis monitoring recommendations, with collection of clinical, biological and ultrasonography data at each visit. It was recommended that the ultrasonography of any given patient be performed at the same center by an experienced operator. All operators completed reports including mentions of the presence or absence of focal liver lesions. If such a lesion was detected, a diagnostic procedure using contrast-enhanced imaging (CT-scan or MRI), serum alpha-fetoprotein assay and/or guided biopsy was performed according to the 2005 American Association for the Study of Liver Diseases Guidelines [1], updated in 2011 [2]. Diagnoses of HCC were thus established by either histological examination performed by an experienced pathologist or probabilistic non-invasive criteria (mainly dynamic imaging revealing early arterial hyper-vascularization and portal washout). Deaths, HCCs, and cirrhosis decompensations occurring during follow-up were extracted from the patient medical records at the participating centers. All information recorded during follow-up was secondarily monitored by a panel of three clinical research associates located at the hepatology department of the Jean Verdier Hospital (AP-HP, Bondy; University Paris 13). All medical diagnoses of events occurring during follow-up were confirmed by two senior hepatologists.

HEPATHER cohort

Follow-up included systematic visits once a year and spontaneous reports for particular events (including death, hepatocellular carcinoma, and cirrhosis decompensation) which were recorded on specific data forms. As HEPATHER was an observational study, decisions about screening for HCC or progression of fibrosis were made by the clinician, but the choices done accorded with French national recommendations, based on the guidelines of the European Association for the Study of the Liver [3]. To retrieve information about patients lost to follow-up, the HEPATHER cohort was merged with the *Système National des Données de Santé* (SNDS) which allowed to collect clinical events of interest until 2018. The SNDS includes the PMSI database (*Programme de Médicalisation des Systèmes d'Information*), which is a comprehensive record of all hospitalizations in France (including public and private hospitalizations) and includes over 30 million public and private admissions per annum [4]. Discharge diagnoses are coded in the PMSI database using the International Classification of Diseases, 10th Revision (ICD-10), which was implemented in France starting in 1997.

Methods used to develop and validate the study prediction model

Identification of the most accurate biomarker for LRC prediction

In the derivation set (CirVir cohort), the LRC prognostic ability of different biomarkers was evaluated with time-dependent ROC curves using the R package `timeROC` [5, 6]. A higher area under the ROC curve at a given time indicates a better ability to discriminate between individuals experiencing the event versus those not experiencing the event prior to that time. Tested biomarkers were AST, ALT, platelets, APRI and FIB4. The most accurate among them was kept as a candidate biomarker for further analyses.

Development of joint latent class models

Joint modelling was used to develop our dynamic predictive tools in the derivation set (CirVir cohort). This approach studies the occurrence of LRCs and the trajectory of a candidate biomarker jointly. To capture heterogeneity at the population level, we employed the Joint Latent Class Models (JLCMs) using the R package `lcmm` [7]. A JLCM includes 3 sub-models: a Cox model (survival part of the model) to study the risk of LRC, a linear mixed effects model (longitudinal part of the JLCM) to identify the trajectories of the biomarker of interest, and a multinomial logistic model that indicates the probability of being in a latent class. The interest of such dynamic modelling is to adjust the risk of LRC over time thanks to repeated measurements. Indeed, a significant evolution of the biomarker of interest during the follow-up results in a switch to another latent class, allowing the refinement of the prognosis estimation. The strategy for constructing our JLCM is to first select the relevant variables for the survival part, then for the longitudinal part, and finally to join the two to obtain the final JLCM.

Survival analysis A Cox proportional risk analysis with backward elimination was performed to select the LRC predictors. The tested baseline variables were gender, arterial hypertension, dyslipidemia, diabetes, BMI, Gamma-GT, bilirubin, albumin, prothrombin time, and FIB4.

Longitudinal analysis Linear mixed modelling was performed to explain the different trajectories of the candidate biomarker according to polynomial functions of time. The adjustment variables tested were gender, BMI, diabetes, arterial hypertension, dyslipidemia, Gamma-GT, bilirubin, albumin, and prothrombin time. Backward elimination was performed to select the adjustment variables.

Joint analysis We then jointly analyzed the two preselected models (survival and longitudinal) and only significant variables were retained in the final JLCM. As our aim was to develop predictive tools, we used the Bayesian Information Criterion (BIC) [8] and the Expected Prognostic Observed Cross-Entropy (EPOCE) [9] to determine the optimal number of latent classes. The assessment of conditional independence assumption was performed using a score test [10]. Parameters included in the JLCM were estimated by likelihood maximization and several sets of initial values were used to ensure convergence to the global maximum.

Accuracy of joint latent class models

The accuracy of our prediction model was validated in the prospective « real life » HEP-ATHER cohort. The risk of LRC predicted by our model at different time horizons (one year; two years; three years) was calculated at each visit (baseline visit V0; one-year follow-up visit V1; two-year follow-up visit V2). Patients were categorized into three risk groups ($<1.5\%$ per year; $1.5-3.0\%$ per year; $\geq 3.0\%$ per year), and the curves of cumulative incidence of LRC were plotted for these three groups at each visit. The predicted risk of LRC by our model was then compared with the observed incidence of LRC at the different time horizons. Finally, we also evaluate the calibration of our model using the time-dependent Brier Score [11].

Supplementary results

Development of the study prediction model in the derivation set (CirVir cohort)

Parameters of the joint latent class modelling

For the longitudinal part of the joint model, the FIB4 trajectory was explained according to a quadratic function of time adjusted on gender at the population level, and a linear function of time at the individual level with unstructured class-specific variance-covariance matrix for the random effects. The survival part of the joint model involved a baseline risk function approximated by a class-specific 3-equi-piecewise function and a class-specific effect of diabetic status.

Choice of the optimal number of latent classes

As our interest was focused on prediction, the first selected JLCMs were those with two and three latent classes: they had the lowest EPOCE estimates (**Supplementary Table S3**) and the conditional independence assumption was not rejected (respectively $p=0.540$ and $p=0.586$; **Supplementary Table S3**). It should be noted that this assumption was rejected for the JLCM with one latent class ($p=0.004$; **Supplementary Table S3**), which demonstrated the need to consider repeated FIB4 measurements to predict LRC risk. Finally, joint analysis selected a JLCM with three latent classes since it gave both the lowest BIC and the higher log-likelihood (**Supplementary Table S3**).

Study prediction model in the derivation set

The first profile (class #1), including 61.4% of patients, was characterized by a stable FIB4 trajectory below 2 over time (**Supplementary Figure S1a**), and a low risk of LRC during follow-up (**Supplementary Figure S1b**). For the second profile (class #2, 26.6% of patients), the FIB4 remains stable around 3 after SVR and patients have a higher risk of developing LRC than in the first class. The third profile (class #3, 12.0% of patients) is characterized by very high baseline FIB4 values (around 8), then a decrease over time to stabilize around 3. Despite preserved liver function at baseline (**Supplementary Table S4**), the prognosis was highly different between the three joint classes (**Supplementary Figure S1b**). The worse prognosis of the third class contrasted with the sharp FIB4 improvement observed in this patient subgroup (**Supplementary Figure S1**). The decrease in FIB4 observed during the follow up of class #3 patients was mainly explained by a decrease in serum transaminases linked to the clearance of liver inflammation following SVR (**Supplementary Figure S2**). In contrast, despite a slight and progressive increase over time, their platelet count remained low below 150 G/L in many cases suggesting the persistence of significant portal hypertension and thus severe underlying liver disease.

Supplementary Tables

Supplementary Table S1: Comparison of baseline characteristics between patients of the derivation set (CirVir cohort) who experienced liver-related complications during their follow-up and those who did not

	Liver-related complications during follow-up			
	All (n = 695)	No (n = 645)	Yes (n = 50)	p ^a
Age (years)	57.4 (51.0-64.6)	57.5 (51.1-64.6)	55.8 (50.8-65.2)	0.434
Male sex (%)	456 (65.4)	424 (65.5)	32 (64.0)	0.878
BMI (kg/m ²)	25.7 (22.9-29.1)	25.5 (22.9-29.0)	28.7 (24.9-32.0)	0.004
Diabetes (%)	123 (17.7)	105 (16.2)	18 (36.0)	0.002
Arterial hypertension (%)	197 (28.3)	180 (27.8)	17 (34.0)	0.415
Dyslipidemia (%)	48 (6.9)	45 (7.0)	3 (6.0)	1.000
Prothrombin time (%)	90 (81-99)	91 (82-100)	79 (71-87)	<0.001
Bilirubin (μmol/L)	10.0 (6.8-14.0)	9.3 (6.6-13.9)	13.5 (8.0-20.0)	0.002
Albumin (g/L)	43.0 (40.0-46.0)	43.2 (40.0-46.0)	41.4 (37.0-43.6)	0.004
Platelets (G/L)	154 (107-200)	156 (112-202)	102 (62-177)	<0.001
AST (IU/L)	30 (24-39)	30 (24-39)	39 (31-55)	<0.001
ALT (IU/L)	28 (21-40)	28 (21-40)	36 (26-45)	0.006
GGT (IU/L)	44 (27-76)	42 (27-70)	89 (47-167)	<0.001
APRI	0.52 (0.32-0.86)	0.50 (0.32-0.80)	1.07 (0.47-1.78)	<0.001
FIB4	2.24 (1.47-3.54)	2.15 (1.44-3.32)	4.23 (2.14-6.07)	<0.001

^a by Fisher's exact test for categorical variables, Mann-Whitney for continuous variables.

Continuous variables are expressed as median with 1st and 3rd quartiles into bracket, and categorical data as number of patients with frequency (%) into bracket.

BMI: Body Mass Index; AST: ASpartate aminoTransferase; ALT: ALanine aminoTransferase; GGT: Gamma-GT.

Supplementary Table S2: Comparison of baseline characteristics between patients of the validation set (HEPATHER cohort) who experienced Liver-related complications during their follow-up and those who did not

	Liver-related complications during follow-up			
	All (n = 7,064)	No (n = 6,791)	Yes (n = 273)	P ^a
Age (years)	58.2 (52.3-66.3)	58.1 (52.2-66.2)	60.2 (54.7-69.0)	<0.001
Male sex (%)	3790 (53.7)	3600 (53.0)	190 (69.6)	<0.001
BMI (kg/m ²)	25.0 (22.4-28.0)	25.0 (22.4-28.0)	26.0 (23.3-29.0)	0.002
Diabetes (%)	893 (12.6)	830 (12.2)	63 (23.1)	<0.001
Arterial hypertension (%)	2281 (32.3)	2154 (31.7)	127 (46.5)	<0.001
Dyslipidemia (%)	100 (1.4)	93 (1.4)	7 (2.6)	0.110
Prothrombin time (%)	95 (87-100)	96 (87-100)	85 (75-94)	<0.001
Bilirubin (μmol/L)	10.0 (7.0-14.1)	10.0 (7.0-14.0)	14.5 (9.0-20.6)	<0.001
Albumin (g/L)	43.0 (40.0-45.4)	43.0 (40.4-45.7)	42.0 (37.1-44.0)	<0.001
Platelets (G/L)	202 (160-245)	204 (164-246)	133 (91-179)	<0.001
AST (IU/L)	24 (19-29)	23 (19-29)	30 (24-40)	<0.001
ALT (IU/L)	20 (15-28)	20 (15-28)	26 (19-37)	<0.001
GGT (IU/L)	25 (17-42)	24 (17-40)	58 (28-102)	<0.001
APRI	0.29 (0.21-0.42)	0.29 (0.21-0.41)	0.62 (0.35-1.08)	<0.001
FIB4	1.55 (1.11-2.18)	1.53 (1.10-2.13)	2.72 (1.85-4.87)	<0.001

^a by Fisher's exact test for categorical variables, Mann-Whitney for continuous variables.

Continuous variables are expressed as median with 1st and 3rd quartiles into bracket, and categorical data as number of patients with frequency (%) into bracket.

BMI: Body Mass Index; AST: ASpartate aminoTransferase; ALT: ALanine aminoTransferase; GGT: Gamma-GT.

Supplementary Table S3: Selection of the number of latent classes for the study prediction model in the derivation set (CirVir cohort)

Four joint latent class models were fitted with numbers of latent classes varying from 1 to 4. This table provides the maximum log-likelihood, the BIC, the posterior proportion of each latent class (%class), the assessment of conditional independence assumption (p-value), and the cross-validated EPOCE estimate computed at times $t = 1, 3$ and 5 years. The maximum log-likelihood is a criterion to be maximized, while the BIC and the EPOCE are criteria to be minimized in the choice of the most appropriate model.

Number of latent classes	Log-likelihood	BIC	p ^a	%class					EPOCE				
				#1	#2	#3	#4	t = 1 year	t = 3 years	t = 5 years			
1	-7431.780	14935.54	0.004	100	-	-	-	0.371	0.414	0.348			
2	-6584.836	13294.01	0.540	90.9	9.1	-	-	0.335	0.366	0.307			
3	-6552.230	13281.15	0.586	61.4	26.6	12.0	-	0.330	0.364	0.305			
4	No model convergence												

^a by a score test [10].

∞

Supplementary Table S4: Patient characteristics as a function of the three latent classes identified by the joint latent class model in the derivation set (CirVir cohort)

	Latent class		
	#1 (n = 427)	#2 (n = 185)	#3 (n = 83)
Baseline:			
Age (years)	56.7 (50.1-63.0)	61.0 (55.7-69.6)	58.7 (52.9-69.4)
Prothrombin time (%)	94 (85-100)	86 (77-97)	81 (68-87)
Bilirubin ($\mu\text{mol/L}$)	9.0 (6.1-12.0)	11.5 (8.0-17.0)	16.5 (10.0-21.6)
Albumin (g/L)	44.0 (41.7-46.8)	42.5 (40.0-45.0)	39.7 (36.2-42.0)
Platelets (G/L)	197 (159-228)	125 (104-161)	74 (52-93)
AST (IU/L)	26 (21-32)	32 (26-40)	41 (33-58)
ALT (IU/L)	26 (19-38)	28 (20-38)	29 (23-45)
GGT (IU/L)	38 (24-60)	43 (26-73)	62 (44-161)
APRI	0.33 (0.25-0.46)	0.64 (0.49-0.86)	1.48 (1.02-2.24)
FIB4	1.46 (1.13-1.96)	3.15 (2.58-3.84)	6.20 (5.14-8.62)
LRC during follow-up (%):			
HCC	7 (1.6)	10 (5.4)	8 (9.6)
Cirrhosis decompensation	6 (1.4)	6 (3.2)	13 (15.7)

Continuous variables are expressed as median with 1st and 3rd quartiles into bracket, and categorical data as number of patients with frequency (%) into bracket.

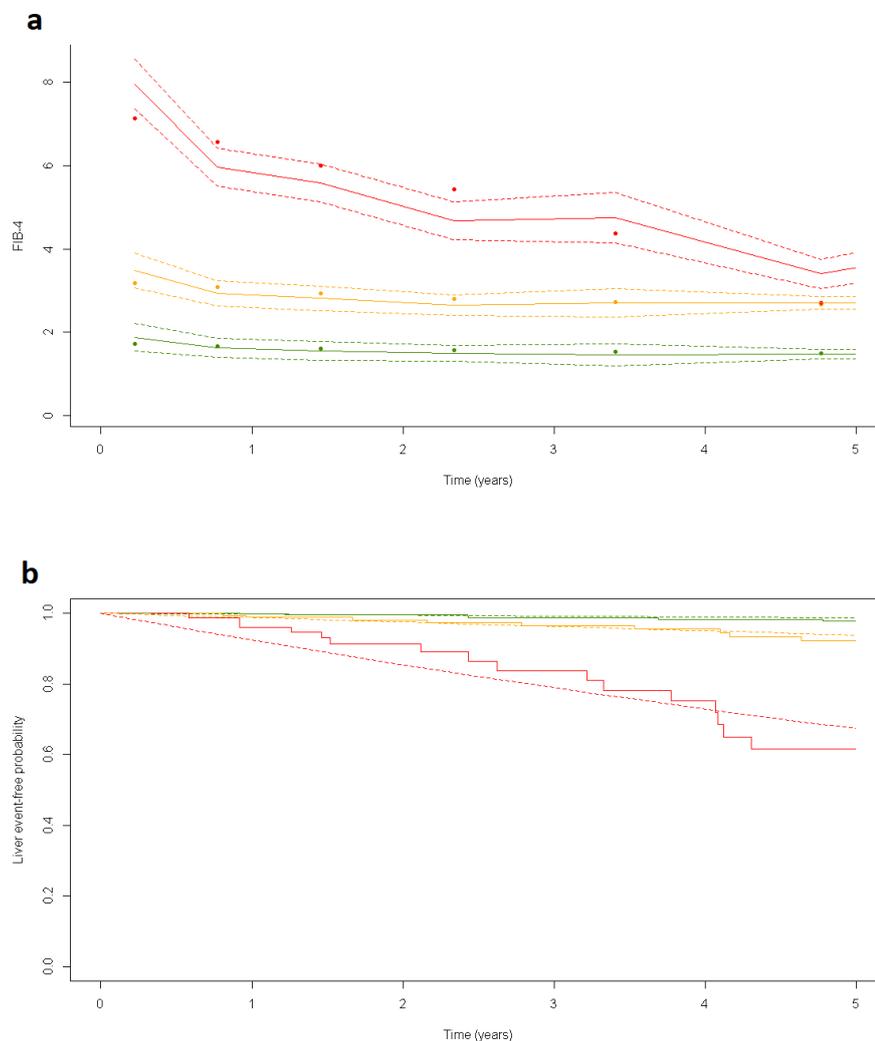
AST: ASpartate aminoTransferase; ALT: ALanine aminoTransferase; GGT: Gamma-GT; HCC: Hepa-toCellular Carcinoma.

Supplementary Figures

Supplementary Figure S1: Joint latent class model developed in the derivation set (CirVir cohort)

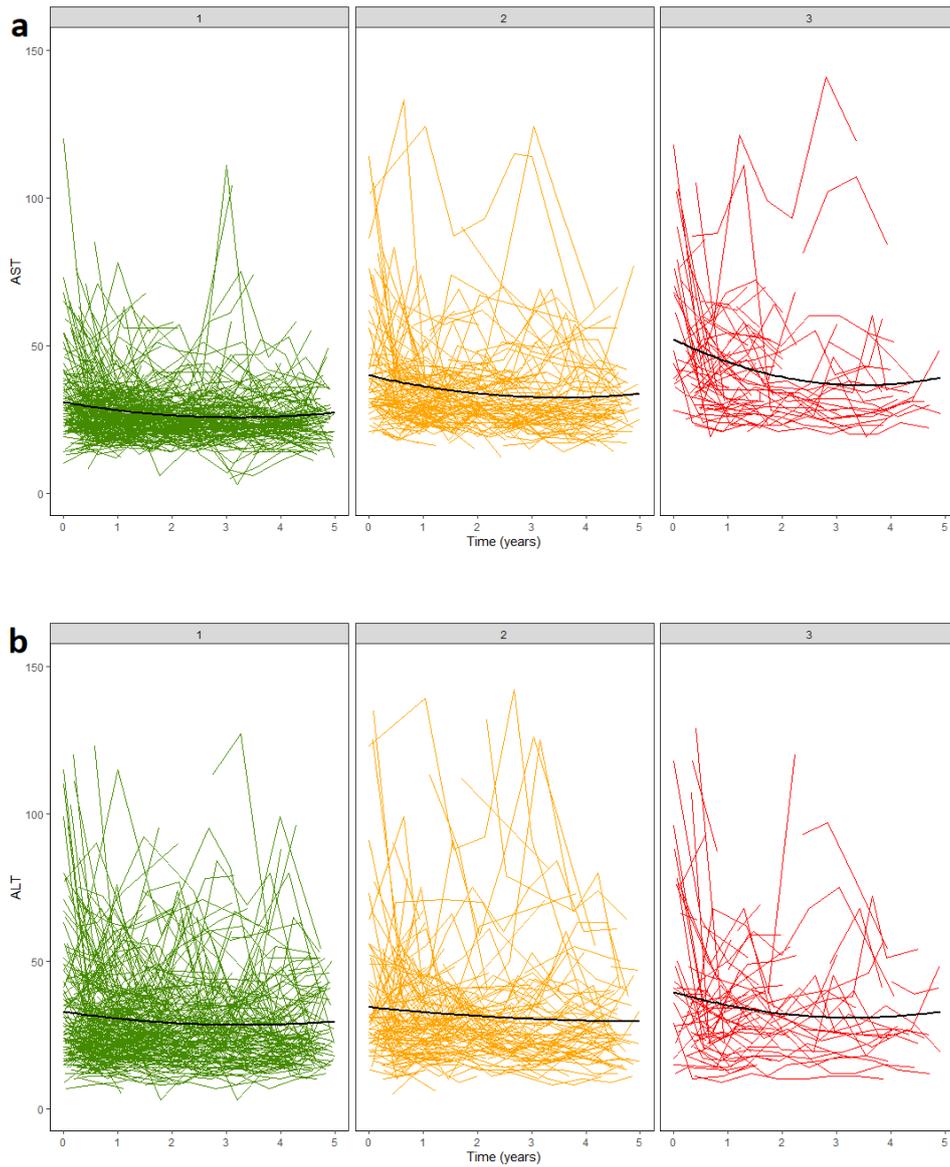
Three patient profiles were identified (green: latent class #1; orange: latent class #2; red: latent class #3).

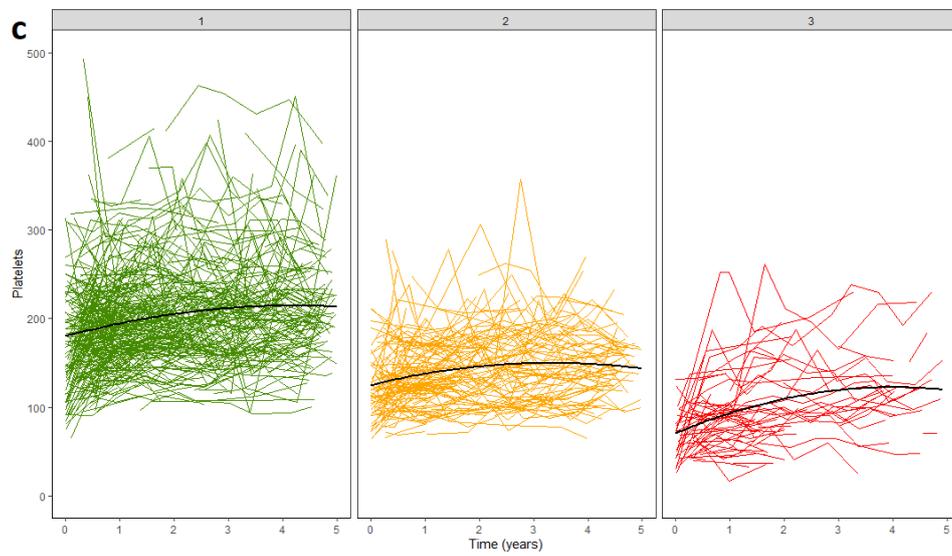
Panel a: Weighted observations and weighted mean of subject-specific predictions from the joint model with three latent classes (● Predicted; — Observed; - - - Confidence Interval); **Panel b:** Survival curves (— Observed curves; - - - Predicted curves)



Supplementary Figure S2: Evolution of serum AST (panel a), ALT (panel b), and platelet count (panel c) during the follow-up in the derivation set (CirVir) cohort, as a function of the latent classes of the joint model

Green lines: patients of class#1; orange lines: patients of class#2; red lines: patients of class#3





References

- [1] Bruix J, Sherman M. Management of hepatocellular carcinoma. *Hepatology* 2005; 42(5): 1208–1236.
- [2] Bruix J, Sherman M. Management of hepatocellular carcinoma: an update. *Hepatology* 2011; 53(3): 1020.
- [3] EASL . EASL recommendations on treatment of hepatitis C 2016. *Journal of Hepatology* 2017; 66(1): 153–194.
- [4] Moulis G, Lapeyre-Mestre M, Palmaro A, Pugnet G, Montastruc JL, Sailler L. French health insurance databases: what interest for medical research?. *La Revue de Médecine Interne* 2015; 36(6): 411–417.
- [5] Blanche P, Dartigues JF, Jacqmin-Gadda H. Estimating and comparing time-dependent areas under receiver operating characteristic curves for censored event times with competing risks. *Statistics in Medicine* 2013; 32(30): 5381–5397.
- [6] Blanche P. Package « timeROC ». <https://cran.r-project.org/web/packages/timeROC/timeROC.pdf>; 2019.
- [7] Proust-Lima C, Philipps V, Lique B. Estimation of extended mixed models using latent classes and latent processes: the R package lcmm. *arXiv preprint arXiv:1503.00890* 2015.
- [8] Schwarz G. Estimating the dimension of a model. *The Annals of Statistics* 1978: 461–464.
- [9] Commenges D, Lique B, Proust-Lima C. Choice of prognostic estimators in joint models by estimating differences of expected conditional Kullback–Leibler risks. *Biometrics* 2012; 68(2): 380–387.
- [10] Jacqmin-Gadda H, Proust-Lima C, Taylor JM, Commenges D. Score test for conditional independence between longitudinal outcome and time to event given the classes in the joint latent class model. *Biometrics* 2010; 66(1): 11–19.
- [11] Graf E, Schmoor C, Sauerbrei W, Schumacher M. Assessment and comparison of prognostic classification schemes for survival data. *Statistics in medicine* 1999; 18(17-18): 2529–2545.

COMPARAISON DE PLUSIEURS CRITÈRES DE PERFORMANCE POUR PRÉDICTIONS DYNAMIQUES

Grâce aux progrès technologiques, de nouveaux biomarqueurs ont pu être mesurés et sont devenus exploitables. En outre, les capacités de stockage des ordinateurs ont également augmenté. L'objectif de nombreuses études cliniques est de trouver des biomarqueurs ayant de fortes capacités pronostiques parmi tous ceux disponibles. En effet, de bons biomarqueurs pronostiques peuvent aider à un diagnostic précoce de la maladie, à identifier de nouvelles cibles pour le développement de médicaments, ou à déterminer le moment optimal pour une intervention.

La plupart des modèles pronostiques sont statiques : ils utilisent des informations à l'inclusion pour faire des prédictions. Cependant, le développement de la modélisation dite dynamique ces dernières années a permis de faire des prédictions de risque basées sur toute l'information disponible durant le suivi du patient. Ces prédictions dynamiques sont mises à jour à chaque nouvelle visite du patient. Cela a contribué au développement de stratégies dynamiques dans le domaine de la médecine personnalisée en adaptant le traitement à chaque patient durant son suivi, en planifiant le moment optimal pour leur prochaine visite de dépistage et en agissant de manière précoce.

Notre objectif dans ce chapitre est d'identifier les meilleurs biomarqueurs pronostiques à suivre dans le temps lorsque plusieurs sont disponibles. Pour illustrer cela, nous allons présenter une application clinique à la NAFLD en comparant les capacités pronostiques de quatre biomarqueurs : le Fibroscan qui est une mesure de l'élasticité du foie, et des tests sanguins FIB-4, APRI, et HEPASCOPE. Nous tenterons de déterminer lesquels de ces quatre biomarqueurs doivent être suivis pour établir le meilleur pronostic des complications hépatiques dans la NAFLD.

Pour évaluer les capacités prédictives de ces quatre biomarqueurs, nous utiliserons l’AUC et le Brier Score dynamiques développés par BLANCHE et al. 2015 pour mesurer la qualité d’une prédiction dynamique. Il existe déjà une procédure de comparaison de deux critères de performance basée sur un intervalle de confiance à un temps landmark ou sur une bande de confiance sur un ensemble de temps landmark (BLANCHE et al. 2015). Cependant, une des limites de ce procédé est que seuls deux critères peuvent être comparés l’un à l’autre. Ce chapitre présente alors une nouvelle procédure de comparaison permettant de comparer plus de deux critères de performance entre eux. Ces résultats sont en cours de soumission.

5.1 Article scientifique

Résumé Avec le développement de la médecine personnalisée, l’étude du pronostic individuel apparaît comme un enjeu scientifique contemporain majeur. Les modèles dynamiques sont particulièrement bien adaptés à de telles études en permettant de prendre en compte certains changements potentiels dans le suivi. Cela permet notamment d’obtenir des prédictions plus précises en mettant à jour les informations disponibles tout au long du suivi du patient. Certains outils mathématiques ont été développés pour quantifier et comparer l’efficacité des prédictions dynamiques en utilisant des versions dynamiques de l’AUC et du Brier Score dans le cadre des risques concurrents. Néanmoins, seules deux prédictions peuvent être comparées. Cela peut être trop restrictif dans un contexte clinique où de plus en plus d’informations peuvent être collectées lors du suivi des patients grâce aux récentes avancées technologiques. Nous proposons ici une nouvelle procédure permettant des comparaisons multiples et basée sur l’AUC et le Brier Score dynamiques. Les performances de notre procédure ont été évaluées par des simulations. De plus, une application en hépatologie visant à identifier le biomarqueur le plus approprié pour prédire les complications liées au foie chez des patients atteints de NAFLD est présentée. La nouvelle procédure permettra de le sélectionner parmi un ensemble de biomarqueurs candidats tout en contrôlant la probabilité de faire au moins une fausse découverte. Pour conclure, ce travail permet de comparer plus de deux critères de performance dynamiques et est disponible via des fonctions R sur GitHub.

Comparison of multiple performance criteria for dynamic predictions

Clémence Moreau¹, Jérémie Riou^{2,3,†}, Marine Roux^{1,†}

¹UPRES 3859, SFR 4208, HIFIH, Angers University, Angers, France

²UMR INSERM 1066, CNRS 6021, MINT, Angers University, Angers, France

³Methodology and Biostatistics Department, Delegation to Clinical Research and Innovation, Angers University Hospital, Angers, France

[†]These authors contributed equally to this article.

Summary With the development of personalized medicine, the study of individual prognosis appears to be a major contemporary scientific issue. Dynamic models are particularly well adapted to such studies by allowing some potential changes in the follow-up to be taken into account. In particular, this leads to more accurate predictions by updating the available information throughout the patient monitoring. Some mathematical tools have been developed to quantify and compare the effectiveness of dynamic predictions using dynamic versions of the Area Under the ROC Curve (AUC) and the Brier score in the competing risks setting. Nevertheless, only two predictions can be compared. This may be too restrictive in a clinical context where more and more information can be collected during patient follow-up thanks to recent technological advances. Here, a new procedure which allows multiple comparisons and based on dynamic AUC or Brier score is proposed. Performances of our testing procedure were assessed by simulations. Moreover, a motivating application in hepatology will be presented, which aimed at identifying the most appropriate biomarker to predict liver-related complications of patients with liver fibrosis. The new procedure will select it among a set of candidate biomarkers while controlling the probability to make at least one false discovery. Finally, this work allows to compare more than two dynamic predictive accuracies and is available through R functions on GitHub.

Keywords Competing risks, Dynamic AUC, Dynamic Brier score, Dynamic prediction, Multiple tests, Prediction accuracy.

1 Introduction

Nowadays, personalized medicine has become a hotspot in the medical approach^[1], which gives a central place to the study of individual prognosis. One of the major challenges in current clinical practice is to identify the factors that are most likely to predict individual risk of disease. This allows clinicians to act early to provide optimal care for the patient, thereby limiting the morbidity in vulnerable patients. Thanks to technological advances, the non-invasive approaches are increasingly being considered and preferred by clinicians. Indeed, they are simple to measure and avoid the side effects of a major invasive procedure.

In practice, the development of non-invasive tools is mainly based on statistical models. The development of so-called dynamic modelling contributes to this end. It allows for dynamic individual risk predictions in the sense that they are updated over time as the patient's clinical profile grows. In particular, joint models^[2, 3] have the advantage of making their predictions by taking into account the entire history of the patient.

Such models should be rigorously evaluated using international guidelines^[4, 5] before being used in clinical practice. Different performance criteria exist to measure the discrimination and calibration of dynamic predictions^[6]. The Area Under the ROC Curve^[7,p. 78] (AUC) is one of the most popular indicators of discrimination and one of the most used in clinical research. It is a recommended tool for measuring the ability of a model to distinguish between patients at risk from those who are not^[8]. The Brier Score^[9] (BS) is another measure of discrimination but also of calibration. It allows us to determine whether the predictions are consistent with the observations. Dynamic versions of the AUC and BS have been developed by Blanche *and al.*^[10] to assess predictive abilities of dynamic predictions in the context of competing risks^[11].

Dynamic AUC is a combination of AUC for competing risks^[12, 13] and AUC for dynamic predictions^[14]. In the same way, the dynamic BS is based on the definitions in the context of competing risks^[15] and dynamic predictions^[16]. Actually, competing events are commonly used in biostatistics. For example, if the event of interest is the complication of the disease, it must be taken into account that the patient may die before. In this case, death would be a competing event. Not to take this into account would be to consider that every patient has a complication before dying, which is not realistic. It is also common in biostatistics to encounter censored data. Indeed, a patient can leave the study if he no longer meets the inclusion criteria, or if he no longer wants to be part of it. The follow-up of a patient can therefore be stopped at a certain point in the study without the patient having experienced the event, leading to bias in the prediction of this event due to missing data. To fix it, Blanche *and al.*^[10] proposed Inverse Probability of Censoring Weighting (IPCW) estimators^[17, 18, 13] of dynamic AUC and BS, able to take into account censored data.

Some tests^[10] are already established to compare two dynamic predictions with each other. The first one, called the *pointwise test*, makes comparison at a given landmark time thanks to a Gaussian confidence interval. While the second called the *simultaneous test*, uses a confidence band calculated by the conditional multiplier central limit theorem to

compare dynamic predictions over all landmark times where they are computed. The null hypothesis of equal performance is rejected as soon as a point on the set of landmark times lies outside the confidence band. As this test does not provide a global p-value, only two dynamic predictions can be compared. However, with recent technological advances, more and more markers are available, leading to the need to compare several markers. Our work addresses this issue. The proposed new test procedure allows the comparison of more than two dynamic predictions while controlling the multiplicity issue. Indeed, when several tests are performed the type one error rate increases. Then, the regulatory agencies (ICH, EMA, FDA) require the Family-Wise Error Rate (FWER) control in confirmatory context, *i.e.* the control of the probability of rejecting at least one null hypothesis incorrectly.

The paper is organised as follows. Section 2 introduces a clinical context in which the developed procedure is needed. Section 3 provides the statistical setting of dynamic predictions, and Section 4 presents the new test procedure to compare multiple performance criteria. Section 5 investigates its performance with simulation studies, and the procedure is applied on motivating example in Section 6. Finally, concluding remarks are given in Section 7.

2 Motivating example

Mainly due to changing eating habits and sedentary lifestyle, obesity or overweight has increased twofold since 1980 to reach nearly a third of the worldwide population^[19]. The World Health Organisation definition of obesity and overweight is an excessive or abnormal fat accumulation that constitutes a health risk^[20]. In this global context, obesity appears to be a major public health problem. One of the risks for the patient is to suffer from Non Alcoholic Fatty Liver Disease (NAFLD). Today, the prevalence of the disease is around 25% and reach more than 80% in obese adults in the general population of Western countries^[21]. This disease can lead to serious liver complications. Of all the liver lesions seen in NAFLD, liver fibrosis is the most important indicator of patient prognosis. Liver biopsy is the gold standard for the assessment of these lesions, but this procedure cannot be widely used in the large NAFLD population due to its invasiveness and associated costs. Several non-invasive tests are now available for the examination of liver fibrosis. Without risk for the patient, these are mainly blood tests and liver elasticity measurements that can easily be repeated during patient follow-up.

In this context, a question of interest for the clinician is to identify the best prognostic markers to follow over time among available non-invasive tests for liver complications. This is key to provide optimal care in order to prevent complications, avoid the risks associated with biopsy and maintain cost-effectiveness. The developed theory was illustrated using the cohort of the hepato-gastroenterology department of the hospital of Angers (France) including 355 patients with a median follow-up of 5.02 [2.55; 7.12] years. In this dataset, several non-invasive tests are available for each patient at several follow-up time points: the Fibroscan^[22], the FIB-4^[23], the APRI^[24] and the HEPASCOPE^[25].

3 Dynamic prediction setting

3.1 Notations

Let T be the time-to-event, C the censoring time, and $\Delta = \mathbb{1}(T \leq C)$ the event indicator. Here, for the sake of simplicity, only two competing events are assumed: $\eta = 1$ for the main event, and $\eta = 2$ for the competing one. Let M be a marker for which $\mathcal{M}_i(s) = \{M_{ij} : 0 \leq t_{ij} \leq s, j = 1, \dots, n_i\}$ represents the information collected for individual i until landmark time s . We consider that the joint distribution of $(T, \eta, \mathcal{M}(s))$ is approximated by a joint model^[2, 3] parametrized by a vector of parameters ξ , estimated by $\hat{\xi}$.

Let $h \in \mathbb{R}^+$ be a fixed prediction horizon. The dynamic prediction^[26] $\pi_i(s, h)$ corresponds to the probability for an individual i to undergo the main event in the time interval $]s, s + h]$ given the information collected until time s . It is defined by

$$\pi_i(s, h) = \mathbb{P}_{\hat{\xi}}(s < T_i \leq s + h, \eta_i = 1 \mid T_i > s, \mathcal{M}_i(s), X_i),$$

with $\mathbb{P}_{\hat{\xi}}$ the probability distribution parametrized by $\hat{\xi}$, and X_i a vector of baseline variables. Moreover, let us denote by $\tilde{T} = \min(T, C)$ the observation time, and $\tilde{\eta} = \Delta\eta$, where $\tilde{\eta}_i = 1$ and $\tilde{\eta}_i = 2$ if the individual i underwent the main event and the competing event, respectively, before being censored; and $\tilde{\eta}_i = 0$ otherwise.

3.2 Performance criteria for dynamic predictions

Let $\left\{ \left(\tilde{T}_i, \Delta_i, \tilde{\eta}_i, \pi_i(\cdot, \cdot) \right), i = 1, \dots, n \right\}$ be an independent and identically distributed (i.i.d) sample of n subjects. Requiring an i.i.d sample implies that the dynamic predictions $\pi_i(\cdot, \cdot)$ should have been fitted on an independent derivation dataset. This part is focused on the prediction of the main event ($\eta = 1$).

Dynamic Area Under the ROC Curve. The dynamic AUC is used to measure the discrimination of dynamic predictions. Let $D_i(s, h) = \mathbb{1}_{(s < T_i \leq s + h, \eta_i = 1)}$ be equal to 1 if the individual i undergoes the main event in the time interval $]s, s + h]$; 0 otherwise. The dynamic AUC of Blanche *and al.*^[10] at the landmark time s and the horizon h is defined by, for all $i, j \in \{1, \dots, n\}$,

$$AUC(s, h) = \mathbb{P}(\pi_i(s, h) > \pi_j(s, h) \mid D_i(s, h) = 1, D_j(s, h) = 0, T_i > s, T_j > s).$$

To deal with censoring, Uno *and al.*^[17] and Hung & Chiang^[18] developed IPCW estimators. This type of estimators consists of weighting individuals by the probability of not being censored^[13]. Assuming that the censoring time C is independent of $(T, \eta, \pi(\cdot, \cdot))$, the weights are defined for a subject i , $i \in \{1, \dots, n\}$, as follows:

$$\widehat{W}_i(s, h) = \frac{\mathbb{1}_{(\tilde{T}_i > s + h)}}{\widehat{G}(s + h \mid s)} + \frac{\mathbb{1}_{(s < \tilde{T}_i \leq s + h)} \Delta_i}{\widehat{G}(\tilde{T}_i \mid s)},$$

where $\widehat{G}(u)$ is the Kaplan-Meier estimator^[13] of $\mathbb{P}(C > u)$, and $\widehat{G}(u \mid s) = \widehat{G}(u) / \widehat{G}(s)$ the probability of not being censored at time u conditionally on being uncensored at time

s. Based on these weights, Blanche *and al.*^[10] propose to estimate the $AUC(s, h)$ by

$$\widehat{AUC}(s, h) = \frac{\sum_{i=1}^n \sum_{j=1}^n \mathbb{1}_{(\pi_i(s, h) > \pi_j(s, h))} \tilde{D}_i(s, h) \left(1 - \tilde{D}_j(s, h)\right) \widehat{W}_i(s, h) \widehat{W}_j(s, h)}{\sum_{i=1}^n \sum_{j=1}^n \tilde{D}_i(s, h) \left(1 - \tilde{D}_j(s, h)\right) \widehat{W}_i(s, h) \widehat{W}_j(s, h)}.$$

Dynamic Brier Score. To measure the discrimination and calibration of dynamic predictions, Blanche *and al.*^[10] also developed the definition of the dynamic BS. Following the definition of the Brier competing risk score of Schoop *and al.*^[16], they defined the dynamic BS at the landmark time s and the horizon h by

$$BS(s, h) = \mathbb{E} \left[(D(s, h) - \pi(s, h))^2 \mid T > s \right],$$

and estimated it by

$$\widehat{BS}(s, h) = \frac{1}{n \widehat{S}_{\tilde{T}}(s)} \sum_{i=1}^n \widehat{W}_i(s, h) \left(\tilde{D}_i(s, h) - \pi_i(s, h) \right)^2,$$

where $\widehat{S}_{\tilde{T}}(s)$ is the estimated probability of observing an individual at risk at time s .

Asymptotic properties. The theory being the same for the dynamic versions of the AUC and the BS, let us denote by $\Theta(\cdot, h)$ either the $AUC(\cdot, h)$ or the $BS(\cdot, h)$ at the h horizon, and $\widehat{\Theta}(\cdot, h)$ the associated IPCW estimator. Let $\tau_0 < \sup \left\{ u : \mathbb{P} \left(\tilde{T} > u \right) > 0 \right\}$, $\tau_1(s) > \inf \left\{ u : \mathbb{P} \left(T \leq s + u, \eta = 1 \mid T > s \right) > 0 \right\}$, and $\tau_2(s) < \sup \left\{ u : \mathbb{P} \left(\tilde{T} > s + u \right) > 0 \right\}$. Here, $[\tau_1(s), \tau_2(s)]$ corresponds to a time interval in which the probability of the main event is non-null, and at least one individual is at risk at $\tau_2(s)$. Let us assume that the censoring time C is independent of $(T, \eta, \pi(\cdot, \cdot))$, then the decomposition of Θ is defined by, for all $s \in [0, \tau_0]$, for all $h \in [\tau_1(s), \tau_2(s)]$,

$$\sqrt{n} \left(\widehat{\Theta}(s, h) - \Theta(s, h) \right) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbb{F} \left(\tilde{T}_i, \tilde{\eta}_i, \pi_i(s, h), s, h \right) + o_p \quad (1)$$

where $\mathbb{F} \left(\tilde{T}_i, \tilde{\eta}_i, \pi_i(s, h), s, h \right)$ is a zero mean influence function defined by Blanche *and al.*^[13, 10], using the theory developed by Van Der Vaart^[27, sec. 12.2].

Equality 1 shows that the estimator $\widehat{\Theta}$ decomposes into a sum of i.i.d. terms. Thus, the central limit theorem gives the convergence of $\widehat{\Theta}$ to a normal centered distribution:

$$\sqrt{n} \left(\widehat{\Theta}(s, h) - \Theta(s, h) \right) \xrightarrow{\mathcal{D}} \mathcal{N} \left(0, \sigma_{s, h}^2 \right),$$

with $\xrightarrow{\mathcal{D}}$ denoting convergence in distribution. A consistent estimator $\widehat{\sigma}_{s, h}^2$ for the variance $\sigma_{s, h}^2$ is given by

$$\widehat{\sigma}_{s, h}^2 = \frac{1}{n} \sum_{i=1}^n \widehat{\mathbb{F}} \left(\tilde{T}_i, \tilde{\eta}_i, \pi_i(s, h), s, h \right)^2$$

where $\widehat{\mathbb{F}} \left(\tilde{T}_i, \tilde{\eta}_i, \pi_i(s, h), s, h \right)$ is an influence function defined in Blanche *and al.*^[10].

4 Comparison of multiple performance criteria

This section presents the developed procedure that compares performances of at least two dynamic predictions. First, for a fixed criteria of interest (AUC or BS), a global comparison will assess whether at least one dynamic prediction is more accurate than the others. Second, post-hoc pairwise comparisons will be performed when appropriate.

4.1 Global comparison

Let $\Theta(s, h)$ be either $AUC(s, h)$ or $BS(s, h)$ available at the landmark time $s \in \mathcal{S} = \{1, \dots, S\}$ and the fixed horizon h , and $\hat{\Theta}(s, h)$ the associated IPCW estimator. Let $\Theta_p(\cdot, h)$, $p \in \mathcal{P} = \{1, \dots, P\}$, be the P quality criteria compared at the fixed horizon h as follows

$$\mathcal{H}_0 : \Theta_1(\cdot, h) = \dots = \Theta_P(\cdot, h) \quad \text{versus} \quad \mathcal{H}_1 : \exists p \in \mathcal{P} \setminus \{\kappa\}, \Theta_p(\cdot, h) \neq \Theta_\kappa(\cdot, h),$$

where Θ_κ is the reference quality criterion determined by

$$\kappa = \arg \max_{x \in \mathcal{P}} \left(\sum_{s \in \mathcal{S}} \sum_{\substack{p \in \mathcal{P} \\ p \neq x}} |\Theta_p(s, h) - \Theta_x(s, h)| \right).$$

Fixing a reference quality criterion ensures the reproducibility of the results. In particular, this choice optimizes the power of the global test and protects against inconsistency between the two steps of the procedure. For a landmark time set \mathcal{S} , a quality criteria set \mathcal{P} and a fixed time horizon h , this null hypothesis is testing through

$$\Psi_{\mathcal{S}, \mathcal{P}, h} = \sum_{s \in \mathcal{S}} \sum_{\substack{p \in \mathcal{P} \\ p \neq \kappa}} \left\{ \frac{\sqrt{n}}{\hat{\sigma}_{s, h}^{p, \kappa}} \left(\hat{\Theta}_p(s, h) - \hat{\Theta}_\kappa(s, h) \right) \right\}^2,$$

where $(\hat{\sigma}_{s, h}^{p, \kappa})^2 = \frac{1}{n} \sum_{i=1}^n \left\{ \mathbb{F} \left(\tilde{T}_i, \tilde{\eta}_i, \pi_p(s, h), s, h \right) - \mathbb{F} \left(\tilde{T}_i, \tilde{\eta}_i, \pi_\kappa(s, h), s, h \right) \right\}^2$, with \mathbb{F} the influence function defined in Web appendix A of Blanche *and al.*^[10].

Theorem 1. *Under \mathcal{H}_0 , for a landmark time set \mathcal{S} , a quality criteria set \mathcal{P} and a fixed time horizon h , the statistical test $\Psi_{\mathcal{S}, \mathcal{P}, h}$ converges to a Gamma distribution*

$$\Psi_{\mathcal{S}, \mathcal{P}, h} \xrightarrow{\mathcal{D}} \Gamma \left(\frac{P-1}{z}; z \right), \quad \text{with } z = 2 \left(1 + \frac{2(P-1) \sum_{k \neq k'}^S \rho_{kk'}}{S(P-1)} \right);$$

where $\rho_{kk'} = \text{Corr}(Z(k, h), Z(k', h))$, calculated by bootstrapping, is the correlation between the different landmark times k and k' , $(k, k') \in \mathcal{S}^2$, $k \neq k'$, of Z defined by

$$Z(k, h) = \sum_{\substack{p \in \mathcal{P} \\ p \neq \kappa}} \left\{ \frac{\sqrt{n}}{\hat{\sigma}_{k, h}^{p, \kappa}} \left(\hat{\Theta}_p(k, h) - \hat{\Theta}_\kappa(k, h) \right) \right\}^2.$$

The proof of this theorem is available in Supporting Information A.1.

Global comparison test. Let \hat{z} be an estimator of z , and $q_{n,\alpha}(\hat{z})$ be the $(1 - \alpha)$ quantile of $\Gamma\left(\frac{P-1}{\hat{z}}; \hat{z}\right)$. The global comparison rejects \mathcal{H}_0 when $\Psi_{S,P,h} \geq q_{n,\alpha}(\hat{z})$, where $q_{n,\alpha}(\hat{z})$ is computed using simulated samples of $\Gamma\left(\frac{P-1}{\hat{z}}; \hat{z}\right)$.

This procedure allows an asymptotic control of the type I error.

Theorem 2. *Assume that \hat{z} is a consistent estimator of z . Then, the global comparison test provides an asymptotic control of the type I error at level α , namely, under the null*

$$\limsup \mathbb{P}(\Psi_{S,P,h} \geq q_{n,\alpha}(\hat{z})) \leq \alpha.$$

The proof of Theorem 2 is given in Supporting Information A.2.

4.2 Post-hoc pairwise comparisons

If, at least one dynamic prediction differs from the reference, pairwise comparisons are performed to identify the differences. The multiplicity issue will be addressed by a multiple testing procedure that controls the Family-Wise Error Rate (FWER), that is the probability of making at least one incorrect rejection. Indeed, when many hypotheses are tested, the latter increases with the number of hypotheses. For each comparison, we finally test for all $p_1, p_2 \in \{1, \dots, P\}^2, p_1 \neq p_2$,

$$\mathcal{H}_0^{p_1 p_2} : \Theta_{p_1}(\cdot, h) = \Theta_{p_2}(\cdot, h) \text{ versus } \mathcal{H}_1^{p_1 p_2} : \Theta_{p_1}(\cdot, h) \neq \Theta_{p_2}(\cdot, h).$$

To control the FWER, several procedures exist. The most naive and easy to use procedure is the well-known Bonferroni correction^[28]. However, it is known to be very conservative when the number of hypotheses to be tested increases or the hypotheses to be tested are correlated. It has been improved by the Bonferroni-Holm method^[29] by applying the closed testing principle. Then, Shaffer^[30] improved the latter by testing only a subset of hypotheses using logical constraints. Recently, in a similar context, this last procedure has demonstrated a good control of the FWER closed to the procedure developed by Blanche *and al.*^[31], that taking into account the correlation between tests. Here, the correlation between tests could not be taken into account due to the unknown multivariate distribution of the tests. Thus, the multiplicity issue will be addressed by the Shaffer procedure.

5 Simulation study

Simulation studies were performed to assess the performance of our procedure. Scenarios of Blanche *and al.*^[10] were replicated for comparison with their procedure when two markers are compared.

5.1 Generated datasets

The datasets were generated with joint latent two classes models considering a latent variable $\gamma = 1, 2$. The class variable followed a Bernoulli distribution of parameter 0.5. Five models were simulated to explain two competing events ($\eta = 1, 2$), where each model was associated with one marker ($l = 1, 2, \dots, 5$) with five years of follow-up assuming a new measurement each year. Moreover, latent class-specific constant risk functions were used for events, and independent censoring was generated with an exponential distribution. Marker trajectories were determined by

$$Y_i^{(l)}(t_{ij}) \Big|_{\gamma_i=g} = \left(\beta_{0g}^{(l)} + b_{i0}^{(l)} \right) + \left(\beta_{1g}^{(l)} + b_{i1}^{(l)} \right) t_{ij} + \epsilon_i^{(l)}(t_{ij})$$

with the latent class $g = 1, 2$, the Gaussian random effects $(b_{i0}^{(l)}, b_{i1}^{(l)}) \sim \mathcal{N}(0, \Sigma_b^{(l)})$ independent of γ_i where $\Sigma_b^{(l)} = \text{diag}(\sigma_{b_0}^{(l)2}, \sigma_{b_1}^{(l)2})$, and $\epsilon_i^{(l)}(t_{ij}) \sim \mathcal{N}(0, \sigma_\epsilon^{(l)2})$ an independent noise. Dynamic predictions were computed (see Supporting Information B) for the five joint latent class models associated with the five markers. The survival parameters were the same under the null and the alternative hypotheses, and the first marker trajectory ($l = 1$) was simulated to be different from the others (see Supporting Information C). Dynamic AUCs were estimated on the set of landmark times $\mathcal{S} = \{0, 1, 2, 3, 4, 5\}$, with a fixed horizon of $h = 5$ years. The average dynamic AUC trajectories are described in Supporting Information D. Simulations were performed on 1000 iterations with sample sizes $n = 500, 1000, 2000$.

5.2 Global procedure

The global procedure was evaluated for the comparison of 2 to 5 markers and compared to simultaneous test of Blanche *and al.*^[10] in the context of two dynamic predictions. The results of the FWER and power of global procedure were summarised in Table 1.

Table 1: Simulation results for the global comparison of dynamic AUCs. Blanche’s test corresponds to the simultaneous test described in Blanche *and al.*^[10].

Number of markers	$n = 500$				$n = 1000$				$n = 2000$			
	2	3	4	5	2	3	4	5	2	3	4	5
FWER												
Global test	4.0	4.4	5.9	6.3	3.2	4.2	5.1	6.0	3.3	4.3	5.3	5.1
Blanche’s test	5.7	-	-	-	4.5	-	-	-	4.5	-	-	-
Power												
Global test	58.3	75.2	85.2	91.8	92.4	98.1	99.8	99.7	99.8	100	100	100
Blanche’s test	69.3	-	-	-	96.8	-	-	-	100	-	-	-

For two dynamic predictions, for small sample size ($n = 500$), the power of the global test was closed to that of Blanche *and al.*^[10], and allowed a control of the type I error in contrast to the latter. Until three markers, the global procedure controlled the FWER.

As expected, when the number of markers increased, more individuals were required to guarantee the FWER control.

5.3 Post-hoc pairwise comparisons

The performances of the pairwise comparisons procedure were assessed through the FWER and the conjunctive power^[32,p. 16]. All results were summarised in Table 2.

Table 2: Simulation results for the post-hoc pairwise comparisons of dynamic AUCs.

Number of markers	$n = 500$			$n = 1000$			$n = 2000$		
	3	4	5	3	4	5	3	4	5
FWER									
Unadjusted	4.1	7.9	16.7	3.2	8.9	14.5	3.1	8.7	15.7
Bonferroni	1.2	0.8	1.4	0.9	1.3	1.6	1.0	1.1	1.8
Bonferroni-holm	2.1	1.1	1.7	2.5	2.0	2.0	3.1	2.4	3.4
Shaffer	3.0	1.6	2.3	3.2	2.5	2.3	3.1	2.4	3.5
Conjunctive power									
Unadjusted	36.0	33.2	22.5	81.6	79.0	73.5	99.6	99.7	99.7
Bonferroni	18.9	11.0	3.5	62.4	46.7	29.3	98.6	95.6	92.0
Bonferroni-holm	23.5	13.8	4.7	68.9	54.0	33.8	99.2	96.8	94.1
Shaffer	31.2	19.9	8.4	79.1	64.5	45.8	99.5	98.5	97.4

All multiple testing procedures provided a FWER control, and Shaffer's outperformed all the others. In particular, it allowed an improvement of the Bonferroni-Holm power. Finally, the power increased with the sample size, but the latter did not impact the control of the FWER.

6 Application to prediction of liver-related risk in NAFLD

As mentioned in Section 2, the NAFLD is a disease where the main prognostic indicator for the patient is the liver fibrosis. Four non-invasive tests could be used to measure it: the Fibroscan which is a measure of liver elasticity and 3 blood scores the FIB-4, the APRI, and the HEPAScore. For this application, the event of interest was hepatic complication which is a composite endpoint including at least one of the following events: ascites, oesophageal variceal rupture, hepatic failure, encephalopathy and hepato-cellular carcinoma. The aim was to determine which of the four biomarkers previously mentioned should be monitored to establish the best prognosis for liver complications. The cohort of the hepato-gastroenterology department of Angers hospital included 355 french NAFLD patients between April 2004 and April 2018, and was split into a derivation set of 236 patients and a validation set of 119 patients. Patients with a history of bariatric surgery, a history of liver complication before baseline, and/or follow-up of less than three months were excluded. In particular, this means that the data is left-truncated. For each visit, patients had a measurement of the four biomarkers.

6.1 Modelisation on derivation set

Patient characteristics at baseline in the derivation set were available in Table 3.

Table 3: Descriptive analysis of variables at baseline in the derivation set, $n = 236$. Variables are described by median[IQR] for quantitative variables, and n(%) for the qualitative ones.

Variables	All patients $n = 236$	No event $n = 213$	Event $n = 23$
Age (years)	58.60 [50.79 ; 64.65]	58.01 [49.93 ; 64.14]	63.46 [59.61 ; 65.72]
Sex (woman)	84 (35.59)	74 (34.74)	13 (56.52)
Treatment			
Antidiabetic	98 (41.53)	81 (38.03)	17 (73.91)
Antihypertensive	139 (59.15)	119 (56.13)	20 (86.96)
Lipid-lowering	97 (41.45)	87 (41.23)	10 (43.48)
Fibroscan (kPa)	8.80 [6.20 ; 13.30]	8.30 [6.10 ; 11.90]	23.30 [12.15 ; 40.15]
FIB-4	1.38 [0.99 ; 2.05]	1.32 [0.93 ; 1.84]	2.48 [1.84 ; 4.45]
APRI	0.44 [0.31 ; 0.69]	0.44 [0.30 ; 0.64]	0.79 [0.56 ; 1.55]
HEPASCORE	0.07 [0.02 ; 0.26]	0.06 [0.02 ; 0.16]	0.66 [0.30 ; 0.98]
Time to event (years)	5.22 [2.66 ; 7.17]	5.42 [2.95 ; 7.28]	1.20 [0.19 ; 6.08]
Number of visits	4 [3 ; 5]	4 [3 ; 5]	2 [1 ; 3]

Time-to-event was calculated from the study entry into the cohort to the occurrence of the liver complication. Patients who did not experience hepatic complication before their last visit were considered event-free. The median follow-up time was 5.22 years with a median number of visits of 4. The patients had a median age of 58.60 years, of which 36% were women. Nearly 40% of the subjects were under antidiabetic treatment, 60% under antihypertensive treatment and 40% under lipid-lowering treatment. Among the 236 patients with NAFLD, 23 experienced a liver complication. These patients appeared to be older than those without events. They were also more likely to be on antidiabetic and antihypertensive treatment, and had higher median values for the four monitored biomarkers (Fibroscan, FIB-4, APRI and HEPASCORE).

To build the dynamic predictions associated with the biomarkers, joint latent class models were fitted in the same way for each biomarker. Models with two latent classes were selected. For each model, the longitudinal trajectory of the biomarker was explained by an intercept and time, both in fixed and random effects. Survival was explained by the biomarker with class-specific Weibull hazard functions.

6.2 Dynamic prediction computation and comparison test procedure on validation set

Based on the joint models fitted in the derivation set, the dynamic predictions were obtained in the validation set using the R package `lcmm`^[33] for different landmark times

$s \in \mathcal{S} = \{0, 0.5, 1, 1.5, 2, 2.5\}$ with a prediction horizon of $h = 3$ years. The choice of a 3-year time horizon was guided by the natural history of fibrosis. Indeed, few years are necessary to develop an hepatic complication. Patient characteristics at baseline in the validation set were available in Table 4. The patients who experienced a liver complication, were again more likely to be on antidiabetic and antihypertensive treatment, and had higher median values for the four monitored biomarkers (Fibroscan, FIB-4, APRI and HEPAScore).

Table 4: Descriptive analysis of variables at baseline in the validation set, $n = 119$. Variables are described by median[IQR] for quantitative variables, and n(%) for the qualitative ones.

Variables	All patients $n = 119$	No event $n = 107$	Event $n = 12$
Age (years)	58.41 [51.38 ; 64.26]	58.41 [51.09 ; 64.26]	58.54 [55.33 ; 61.08]
Sex (woman)	43 (36.13)	38 (35.51)	7 (58.33)
Treatment			
Antidiabetic	50 (42.01)	41 (38.32)	9 (75.00)
Antihypertensive	79 (66.39)	69 (64.49)	10 (83.33)
Lipid-lowering	51 (42.86)	44 (41.12)	7 (58.33)
Fibroscan (kPa)	9.70 [6.40 ; 15.55]	8.90 [6.10 ; 13.00]	30.70 [19.90 ; 36.48]
FIB-4	1.44 [1.10 ; 2.06]	1.37 [1.06 ; 1.93]	2.58 [1.74 ; 4.10]
APRI	0.47 [0.35 ; 0.66]	0.45 [0.35 ; 0.64]	0.67 [0.52 ; 1.62]
HEPAScore	0.07 [0.03 ; 0.16]	0.07 [0.03 ; 0.16]	0.11 [0.01 ; 0.31]
Time to event (years)	4.58 [2.29 ; 6.93]	4.75 [2.75 ; 7.26]	2.37 [0.39 ; 4.13]
Number of visits	3 [2 ; 5]	4 [3 ; 5]	2 [1 ; 4]

The dynamic AUCs were estimated from dynamic predictions based on Fibroscan, FIB-4, APRI, and HEPAScore biomarkers over all landmark times $s \in \mathcal{S}$ years with a 3-year time horizon. The different trajectories of these AUCs are presented in Figure 1.

To determine whether monitoring one biomarker is more prognostic than another, the first global test was performed:

$$\mathcal{H}_0 : AUC_{\text{Fibroscan}}(\cdot, 3) = AUC_{\text{FIB-4}}(\cdot, 3) = AUC_{\text{APRI}}(\cdot, 3) = AUC_{\text{HEPAScore}}(\cdot, 3)$$

versus

$$\mathcal{H}_1 : \exists p \in \{\text{Fibroscan, FIB-4, APRI}\}, AUC_p(\cdot, 3) \neq AUC_{\text{HEPAScore}}(\cdot, 3)$$

where the AUC associated with the HEPAScore dynamic predictions was identified as the reference as described in Section 4.1. The null hypothesis was rejected at the significance level of 5% (p-value < 0.001): at least one of the biomarkers had different prognostic capacities than the others. As differences were found in the global comparison, post-hoc pairwise comparisons were applied. The results are summarised in Table 5.

The prognostic capabilities of HEPAScore are significantly weaker than those of all other biomarkers at a significance level of 5%. No significant difference between Fibroscan,

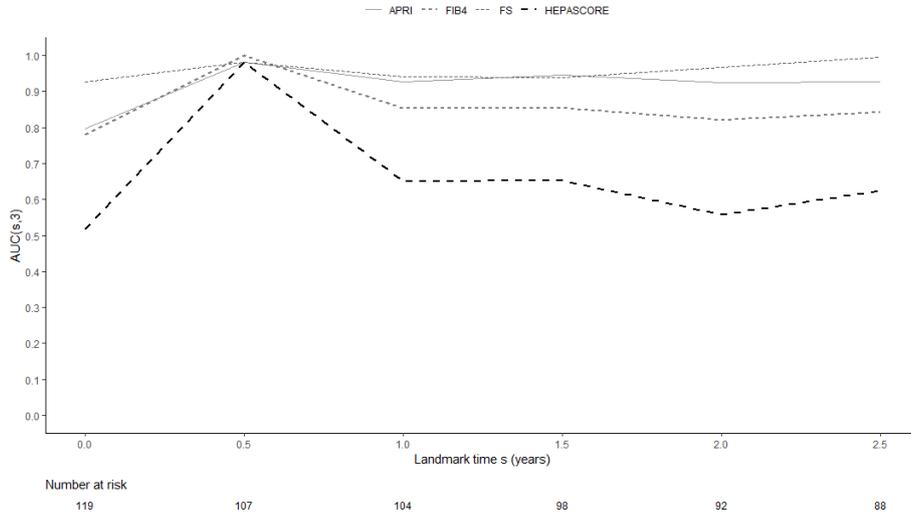


Figure 1: $AUC(s, 3)$ estimated from dynamic predictions based on Fibrosan (FS), FIB-4, APRI, HEPAScore biomarkers over landmark times $\mathcal{S} = \{0, 0.5, 1, 1.5, 2, 2.5\}$ with a prediction horizon of $h = 3$ years.

Table 5: Pairwise comparisons of the four performance criteria associated with Fibrosan, FIB-4, APRI and HEPAScore.

	Fibrosan VS FIB-4	Fibrosan VS APRI	Fibrosan VS HEPAScore	FIB-4 VS APRI	FIB-4 VS HEPAScore	APRI VS HEPAScore
Shaffer p -values	0.218	0.385	0.020	0.273	0.049	0.028

FIB-4 and APRI can be highlighted here: all are relevant prognostic markers of liver complications in NAFLD. The choice of biomarker will therefore depend on the clinician requesting the test: Fibrosan is only available in specialized liver disease centers, while the other two are measured by a simple blood test.

7 Concluding remarks

The proposed methodology allows the comparison of the overall predictive accuracy of at least two dynamic predictions in a prognostic study. The latter is valid in competing risks setting, and also when there is only one type of event. The discrimination and the calibration are compared though the AUC and the Brier score, respectively. When more than two dynamic predictions are compared, our methodology is a two-step procedure. A first p -value results from the global comparison which aims at assessing if at least one of the dynamic predictions has an overall predictive accuracy that differs from the reference

dynamic prediction. The latter was chosen to optimize the power of the procedure, and to avoid problems of inconsistency between the two steps of the procedure. If and only if there is a significant difference at the first step, post-hoc pairwise comparisons are realised to identify the differences. As the joint distribution of tests is unknown, Shaffer's procedure was subsequently used to control the FWER in pairwise comparisons. Indeed, as demonstrated by Blanche *and al.*^[31], this procedure allows to obtain closed performance to a procedure which takes into account both logical constraints described by Westfall^[34] and dependencies between tests.

The performances of our test procedure were investigated by simulation studies. In the context of the comparison of two dynamic predictions, the latter seems to control the type I error even for small sample size, unlike the procedure of Blanche *and al.*^[10], at the price of a small loss in power between both. As expected, the higher the number of markers, the larger the sample size should be to control the type I error. Furthermore, since the asymptotic properties of the AUC and BS estimators require independence between censoring and biomarkers, the performances of our test procedure will be directly impacted by this assumption. According to marker-dependent censoring simulations performed by Blanche *and al.*^[10], the inference procedure for the AUC is rather robust, unlike that for the Brier score. Besides, from a clinical point of view, our procedure allowed to identify the most relevant biomarkers to predict liver-related complication.

It should be noted that our methodology has been applied to compare the overall predictive accuracy of dynamic predictions associated with different biomarkers, but it can also be used to compare different dynamic models. Moreover, our methodology is easily adapted to the case where all pairwise comparisons are not of interest. For example, this may be the case when clinical knowledge is taking into account a priori, leading to an increase in the power of the procedure. Finally, the potential applications of our methodology go far beyond the scope of the hepatology presented in this article. It can be applied as soon as the evaluation of the prognostic impact of a longitudinal marker or a model is of interest. Finally, the R code of our methodology is available on the GitHub https://github.com/ClemenceMoreau/Comparison_multiple_performance_criteria, and the Supporting Information E provides an example for using it.

Acknowledgments

We warmly thank Professor Boursier for sharing with us the database used to illustrate our methodology on an example in hepatology. We also warmly thank Mrs Courcoul for her participation in the development of the R code of our methodology, during her second year internship in ENSAI, Rennes.

Conflict of interest

The authors declare no potential conflict of interests.

Supporting Information

Supporting Information referenced in Sections 5, 6 and 7 contain mathematical proofs of the mentioned theorems, details of the simulation studies, and an application of our procedure with the R code used.

References

- [1] Rizopoulos D, Taylor JM, Van Rosmalen J, Steyerberg EW, Takkenberg JJ. Personalized screening intervals for biomarkers using joint models for longitudinal and survival data. *Biostatistics* 2016; 17(1): 149–164.
- [2] Proust-Lima C, Taylor JM. Development and validation of a dynamic prognostic tool for prostate cancer recurrence using repeated measures of posttreatment PSA: a joint modeling approach. *Biostatistics* 2009; 10(3): 535–549.
- [3] Rizopoulos D. Dynamic predictions and prospective accuracy in joint models for longitudinal and time-to-event data. *Biometrics* 2011; 67(3): 819–829.
- [4] Bossuyt PM, Reitsma JB, Bruns DE, et al. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. *Clinical chemistry* 2015; 61(12): 1446–1452.
- [5] Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *Journal of British Surgery* 2015; 102(3): 148–158.
- [6] Gerds TA, Cai T, Schumacher M. The performance of risk prediction models. *Biometrical Journal: Journal of Mathematical Methods in Biosciences* 2008; 50(4): 457–479.
- [7] Pepe MS, others . *The statistical evaluation of medical tests for classification and prediction*. Medicine . 2003.
- [8] Blanche P, Kattan MW, Gerds TA. The c-index is not proper for the evaluation of year predicted risks. *Biostatistics* 2019; 20(2): 347–357.
- [9] Brier GW. Verification of forecasts expressed in terms of probability. *Monthly weather review* 1950; 78(1): 1–3.
- [10] Blanche P, Proust-Lima C, Loubère L, Berr C, Dartigues JF, Jacqmin-Gadda H. Quantifying and comparing dynamic predictive accuracy of joint models for longitudinal marker and time-to-event in presence of censoring and competing risks. *Biometrics* 2015; 71(1): 102–113.
- [11] Fine JP, Gray RJ. A proportional hazards model for the subdistribution of a competing risk. *Journal of the American statistical association* 1999; 94(446): 496–509.
- [12] Zheng Y, Cai T, Jin Y, Feng Z. Evaluating prognostic accuracy of biomarkers under competing risk. *Biometrics* 2012; 68(2): 388–396.

-
- [13] Blanche P, Dartigues JF, Jacqmin-Gadda H. Estimating and comparing time-dependent areas under receiver operating characteristic curves for censored event times with competing risks. *Statistics in medicine* 2013; 32(30): 5381–5397.
- [14] Parast L, Cheng SC, Cai T. Landmark prediction of long-term survival incorporating short-term event time information. *Journal of the American Statistical Association* 2012; 107(500): 1492–1501.
- [15] Schoop R, Graf E, Schumacher M. Quantifying the predictive performance of prognostic models for censored survival data with time-dependent covariates. *Biometrics* 2008; 64(2): 603–610.
- [16] Schoop R, Beyersmann J, Schumacher M, Binder H. Quantifying the predictive accuracy of time-to-event models in the presence of competing risks. *Biometrical Journal* 2011; 53(1): 88–112.
- [17] Uno H, Cai T, Tian L, Wei LJ. Evaluating prediction rules for t-year survivors with censored regression models. *Journal of the American Statistical Association* 2007; 102(478): 527–537.
- [18] Hung H, Chiang Ct. Optimal composite markers for time-dependent receiver operating characteristic curves with censored survival data. *Scandinavian journal of statistics* 2010; 37(4): 664–679.
- [19] Institute For Health Metrics and Evaluation. Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2015 (GBD 2015) Life Expectancy, All-Cause and Cause-Specific Mortality 1980-2015. 2016.
- [20] World Health Organisation (WHO). Obesity and overweight. Fact sheet. <https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight>; Updated on June 2021, accessed on November 2021.
- [21] Younossi ZM, Golabi P, Avila dL, et al. The global epidemiology of NAFLD and NASH in patients with type 2 diabetes: a systematic review and meta-analysis. *Journal of hepatology* 2019; 71(4): 793–801.
- [22] Foucher J, Chanteloup E, Vergniol J, et al. Diagnosis of cirrhosis by transient elastography (FibroScan): a prospective study. *Gut* 2006; 55(3): 403–408.
- [23] Vallet-Pichard A, Mallet V, Nalpas B, et al. FIB-4: an inexpensive and accurate marker of fibrosis in HCV infection. comparison with liver biopsy and fibrotest. *Hepatology* 2007; 46(1): 32–36.
- [24] Wai CT, Greenson 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(2): 518–526.
- [25] Adams LA, Bulsara M, Rossi E, et al. Hepascore: an accurate validated predictor of liver fibrosis in chronic hepatitis C infection. *Clinical chemistry* 2005; 51(10): 1867–1873.

-
- [26] Proust-Lima C, Blanche P. Dynamic predictions. *Wiley StatsRef: Statistics Reference Online* 2014: 1–6.
- [27] Van derVaart A. *Asymptotic statistics*. Cambridge university press . 1998.
- [28] Bonferroni CE. Il calcolo delle assicurazioni su gruppi di teste. *Studi in onore del professore salvatore ortu carboni* 1935: 13–60.
- [29] Holm S. A simple sequentially rejective multiple test procedure. *Scandinavian journal of statistics* 1979: 65–70.
- [30] Shaffer JP. Modified sequentially rejective multiple test procedures. *Journal of the American Statistical Association* 1986; 81(395): 826–831.
- [31] Blanche P, Dartigues JF, Riou J. A closed max-t test for multiple comparisons of areas under the ROC curve. *Biometrics* 2020.
- [32] Bretz F, Hothorn T, Westfall P. *Multiple comparisons using R*. CRC Press . 2016.
- [33] Proust-Lima C, Philipps V, Liquet B. Estimation of extended mixed models using latent classes and latent processes: the R package lmmm. *arXiv preprint arXiv:1503.00890* 2015.
- [34] Westfall PH. Multiple testing of general contrasts using logical constraints and correlations. *Journal of the American Statistical Association* 1997; 92(437): 299–306.

5.2 Documents complémentaires

Comparison of multiple performance criteria for dynamic predictions: Supporting Information

Clémence Moreau¹, Jérémie Riou^{2,3,†}, Marine Roux^{1,†}

¹UPRES 3859, SFR 4208, HIFIH, Angers University, Angers, France

²UMR INSERM 1066, CNRS 6021, MINT, Angers University, Angers, France

³Methodology and Biostatistics Department, Delegation to Clinical Research and Innovation, Angers University Hospital, Angers, France

[†]These authors contributed equally to this article.

A Proofs of Theorems 1 and 2

A.1 Theorem 1

Theorem 1. Under \mathcal{H}_0 , for a landmark time set \mathcal{S} , a quality criteria set \mathcal{P} and a fixed time horizon h , the statistical test $\Psi_{\mathcal{S},\mathcal{P},h}$ converges to a Gamma distribution

$$\Psi_{\mathcal{S},\mathcal{P},h} \xrightarrow{\mathcal{D}} \Gamma\left(\frac{P-1}{z}; z\right), \text{ with } z = 2 \left(1 + \frac{2(P-1) \sum_{k \neq k'}^{\mathcal{S}} \rho_{kk'}}{S(P-1)}\right);$$

where $\rho_{kk'} = \text{Corr}(Z(k, h), Z(k', h))$, calculated by bootstrapping, is the correlation between the different landmark times k and k' , $(k, k') \in \mathcal{S}^2$, $k \neq k'$, of Z defined by

$$Z(k, h) = \sum_{\substack{p \in \mathcal{P} \\ p \neq \kappa}} \left\{ \frac{\sqrt{n}}{\widehat{\sigma}_{k,h}^{p,\kappa}} \left(\widehat{\Theta}_p(k, h) - \widehat{\Theta}_\kappa(k, h) \right) \right\}^2.$$

Proof. Under the null, for all $s \in \mathcal{S}$, for all $p \in \mathcal{P} \setminus \{\kappa\}$, it follows from

$$\sqrt{n} \left(\widehat{\Theta}(s, h) - \Theta(s, h) \right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma_{s,h}^2)$$

that

$$\frac{\sqrt{n}}{\widehat{\sigma}_{s,h}^{p,\kappa}} \left(\widehat{\Theta}_p(s, h) - \widehat{\Theta}_\kappa(s, h) \right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1),$$

where

$$(\hat{\sigma}_{s,h}^{p,\kappa})^2 = \frac{1}{n} \sum_{i=1}^n \left\{ \mathbb{F} \left(\tilde{T}_i, \tilde{\eta}_i, \pi_p(s, h), s, h \right) - \mathbb{F} \left(\tilde{T}_i, \tilde{\eta}_i, \pi_\kappa(s, h), s, h \right) \right\}^2,$$

with \mathbb{F} the influence function defined in Web appendix A of Blanche *and al.*^[1]. By assuming the independence between the P dynamic predictions, this implies^[2] that

$$\sum_{\substack{p \in \mathcal{P} \\ p \neq \kappa}} \left\{ \frac{\sqrt{n}}{\hat{\sigma}_{s,h}^{p,\kappa}} \left(\hat{\Theta}_p(s, h) - \hat{\Theta}_\kappa(s, h) \right) \right\}^2 \xrightarrow{\mathcal{D}} \chi^2(P-1),$$

where $\chi^2(P-1)$ is a chi-squared distribution with $P-1$ degrees of freedom. Then, Theorem 1 follows from Ferrari^[3, Sec. 2, p. 3]. \square

A.2 Theorem 2

Theorem 2. *Assume that \hat{z} is a consistent estimator of z . Then, the global comparison test provides an asymptotic control of the type I error at level α , namely, under the null*

$$\limsup \mathbb{P}(\Psi_{\mathcal{S}, \mathcal{P}, h} \geq q_{n, \alpha}(\hat{z})) \leq \alpha.$$

Proof. For all $\delta > 0$,

$$\mathbb{P}(\Psi_{\mathcal{S}, \mathcal{P}, h} \geq q_{n, \alpha}(\hat{z})) \leq \mathbb{P}(\Psi_{\mathcal{S}, \mathcal{P}, h} > q_{n, \alpha}(z) - \delta) + \mathbb{P}(q_{n, \alpha}(\hat{z}) \leq q_{n, \alpha}(z) - \delta),$$

and using Theorem 1 and the consistency of \hat{z} , we obtain that

$$\limsup \mathbb{P}(\Psi_{\mathcal{S}, \mathcal{P}, h} \geq q_{n, \alpha}(\hat{z})) \leq \mathbb{P}\left(\Gamma\left(\frac{P-1}{z}; z\right) > q_{n, \alpha}(z) - \delta\right).$$

Finally, the asymptotic control of type I error is obtained by letting $\delta \rightarrow 0$. \square

B Dynamic predictions computation

Let $\mathcal{L}_g(\mathcal{Y}_i(s) | X_i)$ be the multivariate Gaussian density of the vector

$$\mathcal{Y}_i(s) = \{Y_{ij} : 0 \leq t_{ij} \leq s, j = 1, \dots, n_i\}$$

conditionally on X_i and on the latent class $\gamma_i = g$. Let us denote by

$$\lambda_{ikg}(u | X_i) = \lim_{dt \rightarrow 0} \frac{1}{dt} \mathbb{P}(u \leq T_i < u + dt, \eta = k | T_i \geq u, X_i, \gamma_i = g)$$

the class-specific hazard of cause $k = 1, 2$ at time u conditionally on X_i and on the latent class $\gamma_i = g$. Let

$$S_g(s | X_i) = \exp \left[- \int_0^s \{ \lambda_{i1g}(u | X_i) + \lambda_{i2g}(u | X_i) \} du \right]$$

be the overall event free survival function conditionally on X_i and on the latent class $\gamma_i = g$ at time s and

$$F_{1g}(s | X_i) = \int_0^s \{\lambda_{i1g}(u | X_i) S_g(u | X_i)\} du$$

the cumulative incidence of the main event $k = 1$ conditionally on X_i and on the latent class $\gamma_i = g$ at time s . As constant class and cause-specific hazard were selected, it follows that $\lambda_{ikg}(u | X_i) \equiv \lambda_{kg}$,

$$S_g(s | X_i) \equiv S_g(s) = \exp\{-(\lambda_{1g} + \lambda_{2g})s\}$$

and

$$F_{1g}(s | X_i) \equiv F_{1g}(s) = \frac{\lambda_{1g}}{\lambda_{1g} + \lambda_{2g}} [1 + \exp\{-(\lambda_{1g} + \lambda_{2g})s\}].$$

Dynamic predictions calculated using joint latent class models are defined by

$$\begin{aligned} \pi_i(s, h) &= \mathbb{P}_{\xi}(s < T_i \leq s + h, \eta_i = 1 | T_i > s, \mathcal{M}_i(s), X_i) \\ &= \frac{\sum_g \mathbb{P}(\gamma_i = g | X_i) \mathcal{L}_g(\mathcal{Y}_i(s) | X_i) \{F_{1g}(s + h | X_i) - F_{1g}(s | X_i)\}}{\sum_g \mathbb{P}(\gamma_i = g | X_i) \mathcal{L}_g(\mathcal{Y}_i(s) | X_i) S_g(s | X_i)} \end{aligned}$$

with $g = 1, 2$ the latent class.

C Simulation parameters

C.1 Survival parameters

The survival data parameters were generated in the same way under the null and the alternative hypotheses. The events $\eta_i | (\gamma_i = g)$ were generated using a Bernoulli distribution with the parameter

$$\frac{\lambda_{1g}}{\lambda_{1g} + \lambda_{2g}}.$$

The time-to-event $T_i^{(g)}$ conditionally on X_i and on the latent class $\gamma_i = g$ defined by

$$T^{(g)}(u) = \frac{-\log(u)}{\lambda_{1g} + \lambda_{2g}}$$

was simulated by inverting the survival function $u \mapsto S_g(u)$. The independent censoring C was modeled by an exponential distribution of parameter λ_C with

$$\mathbb{E}[C] = \frac{1}{\lambda_C} = 20.$$

The class-specific cause-specific hazard parameters of cause $k = 1, 2$, conditionally on X_i and on the latent class $\gamma_i = g$, were

- Latent class $g = 1$: $\lambda_{11} = \frac{1}{6}$ ($k = 1$) ; $\lambda_{21} = \frac{1}{45}$ ($k = 2$) ;
- Latent class $g = 2$: $\lambda_{12} = \frac{1}{20}$ ($k = 1$) ; $\lambda_{22} = \frac{1}{40}$ ($k = 2$).

C.2 Longitudinal parameters

For longitudinal data Y , the variance parameters for random effects and noise do not change by latent class $g = 1, 2$, marker $l = 1, 2, \dots, 5$, or scenario. Let $\sigma_\epsilon^2 = 1.2$ and $\Sigma_b = \text{diag}(0.25^2, 0.05^2)$. Only the intercept $\beta_{0g}^{(l)}$ and the slope $\beta_{1g}^{(l)}$ change according to the simulation under \mathcal{H}_0 (all markers are simulated in the same way) or \mathcal{H}_1 (the first marker differs from the others). Under the null, with $l = 1, 2, \dots, 5$, these parameters were:

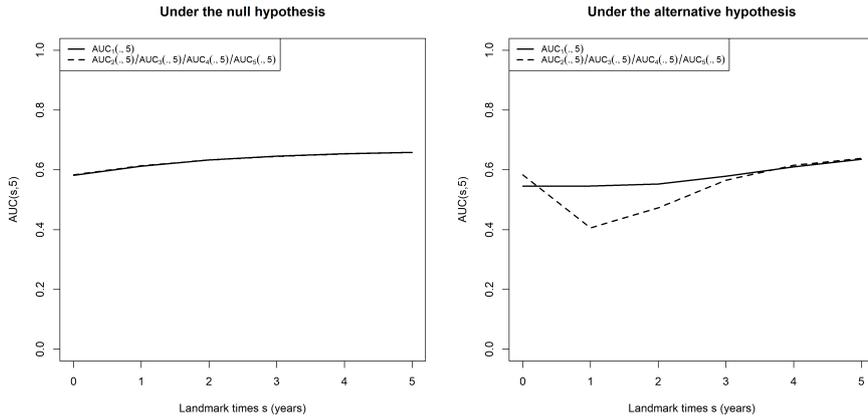
- Latent class $g = 1$: $\beta_{01}^{(l)} = 25$; $\beta_{11}^{(l)} = -0.2$;
- Latent class $g = 2$: $\beta_{02}^{(l)} = 26$; $\beta_{12}^{(l)} = -0.1$.

For the simulations under the alternative hypothesis, the chosen intercept and slope were:

- if $l = 1$,
 - Latent class $g = 1$: $\beta_{01}^{(l)} = 26.5$; $\beta_{11}^{(l)} = -0.5$;
 - Latent class $g = 2$: $\beta_{02}^{(l)} = 26$; $\beta_{12}^{(l)} = -0.1$;
- if $l = 2, 3, 4, 5$,
 - Latent class $g = 1$: $\beta_{01}^{(l)} = 25$; $\beta_{11}^{(l)} = -0.25$;
 - Latent class $g = 2$: $\beta_{02}^{(l)} = 26$; $\beta_{12}^{(l)} = -0.1$.

D Simulated dynamic AUCs trajectories

Figure^{SI} 1 presents the different average trajectories of dynamic AUCs obtained under the null and alternative hypotheses in the simulation studies (1000 iterations, $n = 1000$) in Section 5 of the main manuscript.



Figure^{SI} 1. Average dynamic AUC trajectories under the null hypothesis \mathcal{H}_0 (left) and the alternative hypothesis \mathcal{H}_1 (right).

E Implementation of the comparison procedure in R

An implementation of the comparison procedure using the R functions available on https://github.com/ClemenceMoreau/Comparison_multiple_performance_criteria is presented in this section. For this purpose, we use the validation dataset of 119 NAFLD patients described in Section 6 of the main manuscript where the event studied was the hepatic complication. As a reminder, this set of data included time-to-event data, and also Fibroscan^[4], FIB-4^[5], APRI^[6] and HEPASCOPE^[7] measurements for each patient visit. Dynamic predictions associated with the four biomarkers were computed via the function `dynpred` of the R package `lcmm`^[8]. Table^{SI} 1 is an extract of this dataset and corresponds to the table format taken into account in the R functions used to apply the comparison procedure.

Table^{SI} 1. Dataset used for dynamic prediction comparison in R (sample extracted from the validation dataset used in Section 6 of the main manuscript). This dataset includes the patient identification (« ID »), the time-to-event and event indicator data (« Time-to-event »: numeric type, « Event »: integer type), and the different dynamic predictions computed at $\mathcal{S} = \{0, 0.5, 1, 1.5, 2, 2.5\}$ years (« Landmark times »: numeric type) with a 3-year horizon for Fibroscan (« DP.FS »: numeric type), FIB-4 (« DP.FIB-4 »: numeric type), APRI (« DP.APRI »: numeric type) and HEPASCOPE (« DP.HEP »: numeric type).

ID	Time-to-event	Event	Landmark times	DP.FS	DP.FIB-4	DP.APRI	DP.HEP
1	3.9182	0	0.0	0.0053	0.0474	0.0305	0.0341
1	3.9182	0	0.5	0.0073	0.0415	0.0219	0.0368
1	3.9182	0	1.0	0.0096	0.0389	0.0199	0.0386
1	3.9182	0	1.5	0.0122	0.0372	0.0195	0.0400
1	3.9182	0	2.0	0.0148	0.0358	0.0084	0.0412
1	3.9182	0	2.5	0.0177	0.0346	0.0113	0.0422
2	3.2795	1	0.0	0.2574	0.2152	0.0122	0.0426
2	3.2795	1	0.5	0.1750	0.1809	0.0105	0.0460
2	3.2795	1	1.0	0.1067	0.0164	0.0888	0.0482
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
119	6.4527	0	2.0	0.0179	0.0342	0.1564	0.0327
119	6.4527	0	2.5	0.0212	0.0330	0.1449	0.0334

The R function `Test_comp(tab, horizon, nb_boot, nb_ind_boot)` applies the comparison procedure of dynamic AUCs developed in this manuscript (global comparison, then post-hoc if necessary). It requires:

- a data frame `tab` in the form of Table^{SI} 1 (the columns for ID, Time-to-event, Event, Landmark times and different dynamic predictions must follow the same order as in Table^{SI} 1 but not necessarily the same names),
- the horizon `horizon` (integer type) for which the dynamic predictions have been computed (3 years in this example),

-
- the number of bootstrap samples `nb_boot` (integer type) used to estimate the correlation between the different landmark times,
 - the number of subjects with replacement `nb_ind_boot` (integer type) considered in the bootstrap samples.

To compare the dynamic AUC trajectories of the four dynamic predictions computed for the 3-year horizon in Table^{SI} 1, the following code is executed (1000 bootstrap samples of 119 subjects are used to estimate the correlation).

```
R> Test_comp(tab_valid, 3, 1000, 119)

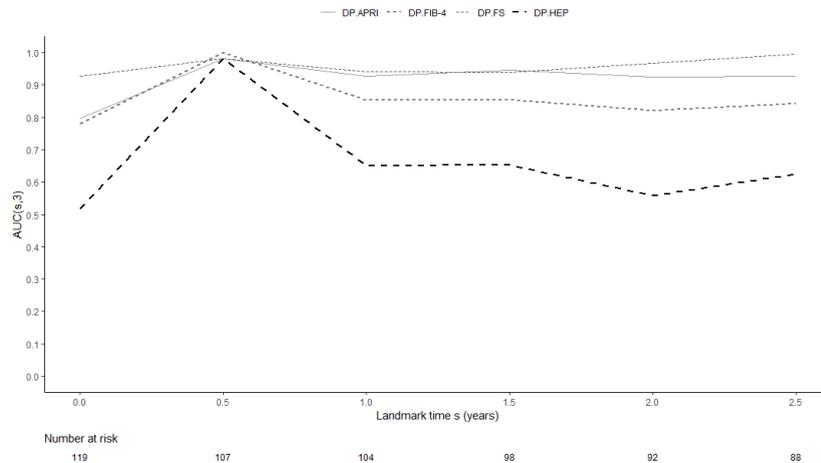
$'Global test'
  Reference          p-val
  DP.HEP      6.0708584e-05

$'Pairwise Shaffer p-values'
 DP.FS_DP.FIB-4   DP.FS_DP.APRI   DP.FS_DP.HEP
  0.2177878       0.3847008       0.01959979
 DP.FIB-4_DP.APRI DP.FIB-4_DP.HEP   DP.APRI_DP.HEP
  0.2727746       0.04895057       0.02750164
```

For the global test, the reference is the dynamic prediction associated to HEPAScore and the p-value is lower than 0.05 (p-value=6.07e-05): the equality of the trajectories is rejected at a 5% level and post-hoc pairwise comparisons are performed. The results indicate a difference between the dynamic predictions associated with HEPAScore and those from the other three biomarkers (Fibroscan, FIB-4 and APRI). If the p-value of the global test had not been significant at a 5% level, the post-hoc tests would not have been performed and only the reference biomarker for the global comparison and the resulting p-value would have been given.

The function `plot_AUC_dyn(tab, horizon)` plots the different dynamic AUCs trajectories according to the different biomarkers and specifies the number of subjects still at risk during the follow-up. The arguments `tab` and `horizon` taken into account are the same ones previously described in the function `Test_comp(tab, horizon, nb_boot, nb_ind_boot)`. Figure^{SI} 2 illustrates the dynamic AUCs trajectories of the dynamic predictions based on Fibroscan, FIB-4, APRI, and HEPAScore for a 3-year time horizon from Table^{SI} 1 using the following code.

```
R> plot_AUC_dyn(tab_valid,3)
```



Figure_{SI} 2. $AUC(s, 3)$ estimated from dynamic predictions based on Fibroscan (DP.FS), FIB-4 (DP.FIB-4), APRI (DP.APRI) and HEPAScore (DP.HEP) over landmark times $\mathcal{S} = \{0, 0.5, 1, 1.5, 2, 2.5\}$ with a prediction horizon of $h = 3$ years.

References

- [1] Blanche P, Proust-Lima C, Loubère L, Berr C, Dartigues JF, Jacqmin-Gadda H. Quantifying and comparing dynamic predictive accuracy of joint models for longitudinal marker and time-to-event in presence of censoring and competing risks. *Biometrics* 2015; 71(1): 102–113.
- [2] Lancaster H. *The Chi-Squared Distribution*. Wiley New York . 1969.
- [3] Ferrari A. A note on sum and difference of correlated chi-squared variables. 2019.
- [4] Foucher J, Chanteloup E, Vergniol J, et al. Diagnosis of cirrhosis by transient elastography (FibroScan): a prospective study. *Gut* 2006; 55(3): 403–408.
- [5] Vallet-Pichard A, Mallet V, Nalpas B, et al. FIB-4: an inexpensive and accurate marker of fibrosis in HCV infection. comparison with liver biopsy and fibrotest. *Hepatology* 2007; 46(1): 32–36.
- [6] Wai CT, Greenson 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(2): 518–526.
- [7] Adams LA, Bulsara M, Rossi E, et al. Hepascore: an accurate validated predictor of liver fibrosis in chronic hepatitis C infection. *Clinical chemistry* 2005; 51(10): 1867–1873.
- [8] Proust-Lima C, Philipps V, Lique B. Estimation of extended mixed models using latent classes and latent processes: the R package lcmm. *arXiv preprint arXiv:1503.00890* 2015.

DISCUSSION

L'approche soutenue durant cette thèse correspond à un concept émergeant dans le domaine de l'hépatologie : il consiste à ne plus poser un diagnostic selon les lésions du foie, mais plutôt de prédire un risque de complications liées à ce dernier pour les patients ayant une fibrose hépatique. Pour cela, de nombreux tests non-invasifs corrélés à la fibrose, incluant biomarqueurs sanguins et élastographiques, peuvent être utilisés comme des marqueurs pronostiques. Contrairement à la biopsie hépatique qui est la référence pour l'évaluation histologique du foie, ils sont facilement répétables dans le temps. Ces tests non-invasifs permettent ainsi d'actualiser le pronostic du patient à chaque nouvelle visite du suivi. L'objectif de cette thèse était d'apporter des réponses statistiques à ce changement de paradigme clinique dans la prise en charge des patients atteints de pathologies hépatiques chroniques.

6.1 Résumés et perspectives des travaux de thèse

Après avoir présenté plus en détails le contexte clinique de cette thèse (voir Chapitre 1) ainsi que les différents outils statistiques auxquels nous avons fait appel (voir Chapitre 2), le Chapitre 3 a permis de comparer les capacités pronostiques de la biopsie face à celles d'un test sanguin (le FIB-4) et d'un test élastographique (le Fibroscan) au sein de la NAFLD. Les résultats obtenus ont montré que le Fibroscan possède des capacités à prédire les complications hépatiques similaires à la biopsie du foie, se montrant ainsi comme une bonne alternative à celle-ci. Nos résultats fournissent également une validation pronostique d'un algorithme séquentiel composé des tests FIB-4 et Fibroscan. Les patients avec un FIB-4 inférieur à 1,30 et ceux avec un FIB-4 supérieur ou égal 1,30 associé à un Fibroscan inférieur à 8,0 kPa ont un excellent pronostic. Au contraire, ceux avec un FIB-4 supérieur ou égal 1,30 avec un Fibroscan supérieur ou égal 8,0 kPa ont un risque plus fort de développer une complication hépatique. Finalement, les tests non-invasifs FIB-4 et Fibroscan permettent une prédiction précise de ces complications chez les patients

atteints de NAFLD : ils pourraient être utilisés à la place de la biopsie hépatique pour l'identification des patients à risque nécessitant une prise en charge spécialisée.

Perspectives du Chapitre 3 Une extension de ce travail pourrait être d'étudier les capacités pronostiques du Fibroscan et du FIB-4 mais cette fois-ci en longitudinal. Cela consisterait à regarder si l'évolution de l'un des deux au cours du temps serait corrélée à l'apparition de complications hépatiques. Le développement d'un modèle longitudinal prédictif comme utilisé lors du Chapitre 4 pourrait également être envisagé. Ainsi, le pronostic du patient ne serait plus établi selon des valeurs ponctuelles du FIB-4 et du Fibroscan comme effectué lors du Chapitre 3, mais plutôt selon toutes les données recueillies durant le suivi.

Dans le Chapitre 4, nous avons développé un modèle conjoint à classes latentes permettant de prédire individuellement le risque de complications hépatiques chez des patients atteints par le VHC. Même si l'obtention d'une RVS chez ces patients est associée à une diminution du risque de complications hépatiques, un risque résiduel subsiste chez les patients cirrhotiques. Le principal challenge pour les médecins est alors d'identifier avec précision ces individus à risque chez qui une prise en charge spécialisée doit être maintenue pour prévenir le développement des complications liées au foie. Le modèle développé inclut le sexe, ainsi que les mesures répétées du FIB-4 et du statut diabétique du patient. Il a été validé dans une grande cohorte prospective de la « vie réelle » comprenant plus de 7000 patients. Un modèle comme celui-ci améliore la prise en charge individuelle des patients infectés par le VHC : il permet d'identifier les malades pour lesquels la surveillance peut être interrompue ou, à l'inverse, ceux où elle doit être renforcée.

Perspectives du Chapitre 4 Dans un premier temps, un second modèle prenant en compte l'élastographie hépatique (comme le Fibroscan par exemple) pourrait être développé. Des données allant dans ce sens doivent être collectées à différents instants du suivi des patients. Par ailleurs, il serait intéressant de décomposer l'évènement étudié « complication hépatique » qui correspondait ici à l'apparition d'un CHC et/ou d'une décompensation de cirrhose. Un modèle multi-états conjoint pourrait alors permettre d'étudier les différents passages observés entre les stades de CHC, de décompensation de cirrhose, et de décès par exemple.

Ce type de médecine personnalisée est aujourd'hui un domaine d'intérêt majeur pour la prise en charge des patients. En identifiant les facteurs les plus susceptibles de prédire

le risque individuel de maladie, les cliniciens peuvent agir précocement et fournir des soins optimaux au patient en limitant ainsi la morbidité chez les patients vulnérables. Pour cela, des modèles dynamiques comme utilisés lors du Chapitre 4 ont été développés afin d'obtenir des prédictions dynamiques individuelles. La méthodologie développée dans le Chapitre 5 permet de comparer les capacités prédictives d'au moins deux prédictions dynamiques dans une étude pronostique. Lorsque plus de deux prédictions de ce type sont comparées, notre méthodologie est une procédure en deux étapes. La première évalue si au moins une des prédictions a une précision prédictive globale qui diffère de la prédiction dynamique de référence (choisie pour optimiser la puissance de la procédure). Seule cette étape est réalisée lors de la comparaison de deux prédictions. Si et seulement s'il y a une différence significative lors de la première étape, des comparaisons post-hoc par paire avec une gestion de la multiplicité des tests sont réalisées pour identifier les différences. Cette nouvelle procédure a été validée à l'aide d'études de simulation, puis a été appliquée sur des données réelles dans le cadre de la NAFLD. À l'aide de celle-ci, nous pouvons ainsi déterminer le meilleur biomarqueur à suivre dans le temps ou le meilleur modèle dynamique au sein d'une étude pronostique.

Perspectives du Chapitre 5 En premier lieu, une matrice de contraste pourrait être intégrée lors de la seconde étape de la procédure. Elle permettrait de sélectionner uniquement les comparaisons par paire ayant un intérêt clinique, augmentant ainsi la puissance de la procédure. Une autre idée pourrait être de découper des fenêtres de temps pour effectuer les différentes comparaisons : cela permettrait par exemple de déterminer plusieurs biomarqueurs d'intérêt pronostique selon la fenêtre observée.

6.2 Nouvelle approche pronostique

Durant cette thèse, nous nous sommes intéressés individuellement à la NAFLD et à l'hépatite C, où les complications observées étaient la cirrhose et/ou le CHC dans chacune d'entre elles. Cependant, ces mêmes complications peuvent survenir chez les patients ayant développés une fibrose du foie, et cela qu'importe la maladie hépatique chronique. Nous pourrions alors nous poser la question du besoin de stratifier les patients selon les pathologies du foie chroniques pour les études pronostiques contre une prise en charge universelle pour ces patients. En effet, la cirrhose est la onzième cause de décès mondiale avec 1250000 morts par an dont 8000 en France (FIALLA, MUCKADELL et TOUBORG

LASSEN 2012). Le CHC compte 830000 décès dans la population mondiale chaque année (9000 en France) représentant la seizième cause de décès. Ces chiffres s'expliquent par une prise en charge trop tardive pour la plupart des patients atteints de maladies chroniques du foie : 75% des cirrhoses sont décompensées au moment de leur diagnostic et 78% des CHC sont au stade palliatif, tous deux étant associés à un mauvais pronostic à court terme (temps de survie médian inférieur à 2 ans). Ces patients sont donc déjà à un stade avancé de la maladie lorsque les cliniciens les rencontrent à l'hôpital. De meilleurs outils pronostiques pour une prise en charge universelle ainsi qu'un dépistage de masse sont alors nécessaires afin de détecter au plus tôt les malades ayant développés une fibrose hépatique, et améliorer leur pronostic.

Comme nous avons pu le dire précédemment, la pratique actuelle consiste à évaluer la fibrose du foie à l'aide d'une biopsie hépatique afin de diagnostiquer la gravité de la maladie. Cependant, cet examen est effectué sur une population sélectionnée et induit des biais (échantillonnage et reproductibilité inter-observateur). Il est donc préférable d'utiliser les tests non-invasifs : ils permettent de mettre en évidence une fibrose hépatique avancée sans passer par l'état lésionnel du foie, et ne présentent pas de risque pour le patient. La Figure 6.1 indique les différentes étapes nécessaires dans ce changement de prise en charge des patients atteints de maladies hépatiques chroniques.

L'expérience en matière de tests non-invasifs des lésions hépatiques a démontré que de simples marqueurs sanguins peuvent identifier un sous-ensemble de patients à risque de fibrose hépatique (EASL 2021). Des études ont montré que le pronostic lié au foie est altéré dans ce sous-groupe (voir Chapitre 3). Cependant, dans la NAFLD par exemple, les tests sanguins simples de fibrose hépatique existants utilisés en première ligne génèrent environ 21% de patients ne présentant pas un stade avancé de la maladie, qui sont orientés inutilement vers l'hépatologue (SRIVASTAVA et al. 2019a). Bien que cette approche se soit avérée rentable financièrement, elle doit être améliorée afin de réduire les coûts inutiles et faire un meilleur usage des ressources de santé. Avec la tension hospitalière et le manque de personnel, il devient crucial de sélectionner uniquement les patients nécessitant un suivi régulier pour éviter de surcharger les services. Le clinicien ne doit pas revoir trop souvent les patients chez lesquels il n'y aura probablement pas de complication, mais ne doit pas non plus en manquer une ou attendre une dégradation trop importante de la fibrose hépatique. De plus, ces mêmes tests non-invasifs présentent des performances limitées dans la population générale (HAGSTRÖM et al. 2020) : une équation de risque

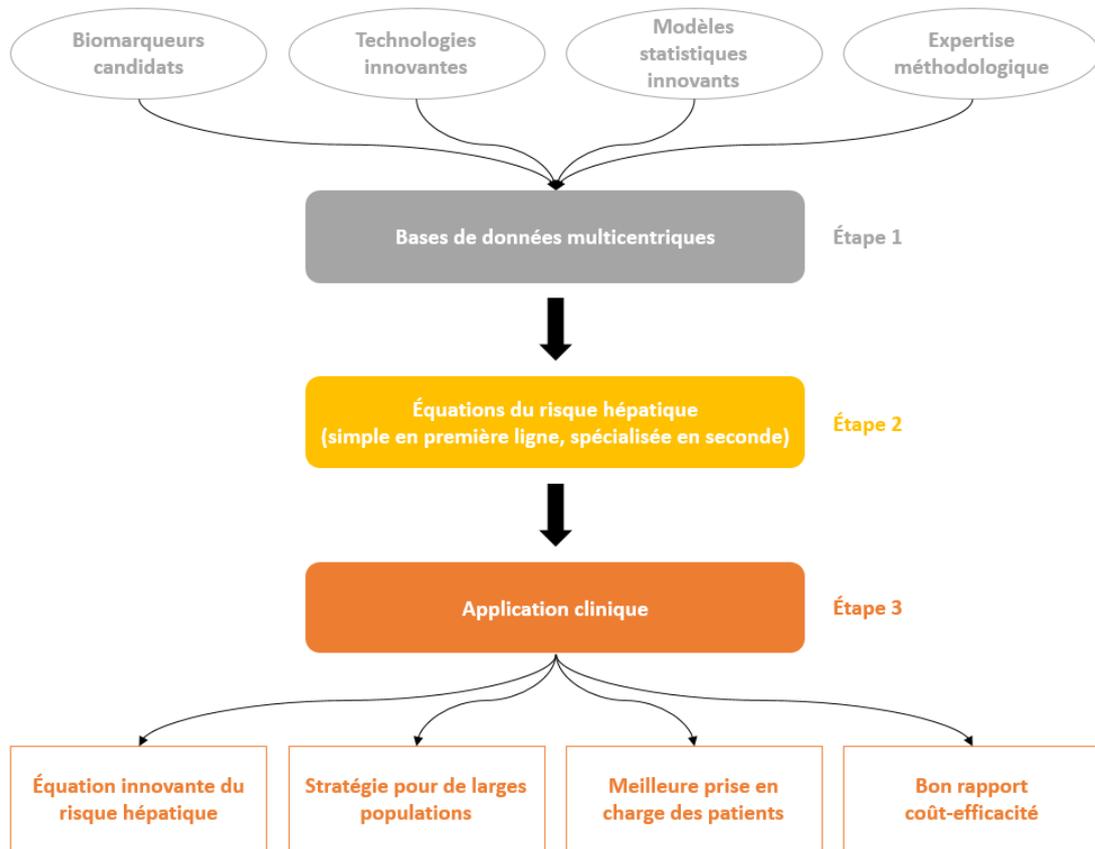


FIGURE 6.1 – Changement de paradigme dans la gestion des maladies hépatiques chroniques

innovante utilisant des marqueurs sanguins de routine serait alors nécessaire dans le but d'identifier les patients les plus à risque de complications hépatiques lors d'un dépistage de masse.

L'un des avantages des tests non-invasifs est qu'ils peuvent être facilement répétés pour surveiller les patients durant le suivi. Un modèle de Cox classique peut être réévalué dans le temps mais il ne prend en compte que les biomarqueurs collectés à chaque visite et non leur évolution depuis le début du suivi. Il semble alors plus pertinent d'utiliser un modèle longitudinal, tel que le modèle conjoint à classes latentes par exemple. Contrairement à un modèle « baseline » répété comme celui de Cox, il permet d'obtenir des prédictions dynamiques basées à la fois sur la trajectoire des biomarqueurs et sur l'occurrence des complications liées au foie. De plus, ce modèle conjoint prend en compte l'hétérogénéité au niveau de la population et ajuste le risque hépatique dans le temps grâce à l'ensemble des

antécédents du patient. D'autres modèles, combinant par exemple une approche landmark étendue à l'historique des biomarqueurs avec des méthodes de *machine learning* adaptées aux données de survie (DEVAUX et al. 2022), peuvent également être utilisés afin d'obtenir ce type de prédictions dynamiques. Une équation de risque développée avec de tels modèles permettrait de prendre en compte l'évolution des biomarqueurs chez les patients sans se restreindre à la valeur sur une seule visite.

Pour développer ce type de modèles, de larges cohortes longitudinales incluant des patients suivis régulièrement à l'aide de tests non-invasifs doivent être disponibles afin d'observer un nombre conséquent de complications hépatiques. Il devient alors essentiel de travailler en réseau en regroupant des bases de données issues de plusieurs centres hospitaliers de différents pays pour obtenir des cohortes multicentriques plus importantes. Des bases de données de « vie réelle » contenant des informations issues de la pratique clinique peuvent aussi être utilisées. L'avantage est qu'elles permettent de faire des analyses sur une large population en ayant tout l'historique médical du patient. Ce type de données étant très hétérogène, il est important de gommer certains biais qui peuvent apparaître à l'aide de différentes méthodes statistiques. En France par exemple, le Système National des Données de Santé (SNDS) est une base nationale créée en 2016 regroupant les données du Système National d'Information Inter-Régimes de l'Assurance Maladie (SNIIRAM), du Programme de Médicalisation des Systèmes d'Information (PMSI) et du Centre d'épidémiologie sur les causes médicales de Décès (CépiDc). Le SNIIRAM contient des informations à propos des soins, consultations, examens, et traitements remboursés d'un patient ; le PMSI décrit les différentes hospitalisations d'un malade, tandis que le CépiDc indique les causes médicales des décès. Ces bases de données importantes constitueraient la socle du changement de la prise en charge des malades chroniques du foie (Étape 1 de la Figure 6.1).

Compte tenu du nombre important de patients à évaluer en pratique clinique, une équation du risque hépatique utilisant des biomarqueurs courants et bon marché serait une solution très pertinente à utiliser par les médecins généralistes dans un premier temps. Un tel type d'équation existe déjà dans le domaine de la cardiologie, pouvant ainsi être appliquée dans la population générale. En effet, le SCORE2 (ESC CARDIOVASC RISK COLLABORATION et SCORE2 WORKING GROUP 2021) est un algorithme qui permet de prédire le risque de maladies cardiovasculaires à un horizon de 10 ans selon l'âge, le statut tabagique, la pression artérielle systolique, et le cholestérol total et HDL du patient.

Ainsi, il conviendrait de développer une équation de risque de routine qui serait utilisée en première ligne, s'intéressant directement au pronostic des évènements hépatiques et non à la fibrose. Par la suite, les patients les plus à risque pourraient ainsi être orientés vers des hépatologues pour une évaluation plus spécifique à l'aide de biomarqueurs sanguins spécialisés ou élastographiques. Une telle stratégie pourrait permettre de diminuer le taux d'orientation inutile vers le spécialiste tout en maintenant la détection des maladies du foie chez les patients atteints de fibrose hépatique.

Un tel outil conduirait à une meilleure organisation du parcours patient : les médecins généralistes initieraient la prise en charge médicale des patients vers les hépatologues (Étape 2 de la Figure 6.1). Une fois appliquées en pratique, ces équations de risque individuel permettraient aux patients les plus à risque d'avoir une prise en charge plus précoce et un meilleur pronostic (Étape 3 de la Figure 6.1). Assurant un bon rapport coût-efficacité, cette stratégie applicable à de larges populations serait innovante dans le domaine de la médecine personnalisée en hépatologie.

BIBLIOGRAPHIE

- AASLD et al. (2015). « Hepatitis C guidance : AASLD-IDSA recommendations for testing, managing, and treating adults infected with hepatitis C virus ». In : *Hepatology* 62.3, p. 932-954.
- ADAMS, Leon et al. (2005). « Hepascore : an accurate validated predictor of liver fibrosis in chronic hepatitis C infection ». In : *Clinical Chemistry* 51.10, p. 1867-1873.
- AKAIKE, Hirotugu (1974). « A new look at the statistical model identification ». In : *IEEE Transactions on Automatic Control* 19.6, p. 716-723.
- ALBERTI, George, Paul ZIMMET et Jonathan SHAW (2005). « The metabolic syndrome—a new worldwide definition ». In : *The Lancet* 366.9491, p. 1059-1062.
- ALLEN, Alina et al. (2018). « Nonalcoholic fatty liver disease incidence and impact on metabolic burden and death : a 20 year-community study ». In : *Hepatology* 67.5, p. 1726-1736.
- ANGULO, Paul (2002). « Nonalcoholic fatty liver disease ». In : *New England Journal of Medicine* 346.16, p. 1221-1231.
- ANGULO, Paul et al. (2007). « The NAFLD fibrosis score : a noninvasive system that identifies liver fibrosis in patients with NAFLD ». In : *Hepatology* 45.4, p. 846-854.
- ANSTEE, Quentin, Giovanni TARGHER et Christopher DAY (2013). « Progression of NAFLD to diabetes mellitus, cardiovascular disease or cirrhosis ». In : *Nature Reviews Gastroenterology & Hepatology* 10.6, p. 330-344.
- BARANOVA, Ancha et al. (2011). « Non-invasive markers for hepatic fibrosis ». In : *BMC Gastroenterology* 11.1, p. 1-15.
- BATTS, Kenneth et Jurgen LUDWIG (1995). « Chronic hepatitis. An update on terminology and reporting. » In : *The American Journal of Surgical Pathology* 19.12, p. 1409-1417.
- BENJAMINI, Yoav et Yosef HOCHBERG (1995). « Controlling the false discovery rate : a practical and powerful approach to multiple testing ». In : *Journal of the Royal statistical society : Series B (Methodological)* 57.1, p. 289-300.
- BETHEA, Emily et al. (2018). « Should we treat acute hepatitis C? A decision and cost-effectiveness analysis ». In : *Hepatology* 67.3, p. 837-846.

-
- BLACH, Sarah et al. (2017). « Global prevalence and genotype distribution of hepatitis C virus infection in 2015 : a modelling study ». In : *The Lancet Gastroenterology & Hepatology* 2.3, p. 161-176.
- BLANCHE, Paul (2013). « Contribution à l'évaluation de capacités pronostiques en présence de données censurées, de risques concurrents et de marqueurs longitudinaux : inférence et applications à la prédiction de la démence ». Thèse de doct. Bordeaux 2.
- BLANCHE, Paul, Jean-François DARTIGUES et Hélène JACQMIN-GADDA (2013a). « Estimating and comparing time-dependent areas under receiver operating characteristic curves for censored event times with competing risks ». In : *Statistics in Medicine* 32.30, p. 5381-5397.
- (2013b). « Review and comparison of ROC curve estimators for a time-dependent outcome with marker-dependent censoring ». In : *Biometrical Journal* 55.5, p. 687-704.
- BLANCHE, Paul, Michael KATTAN et Thomas GERDS (2019). « The c-index is not proper for the evaluation of year predicted risks ». In : *Biostatistics* 20.2, p. 347-357.
- BLANCHE, Paul et al. (2015). « Quantifying and comparing dynamic predictive accuracy of joint models for longitudinal marker and time-to-event in presence of censoring and competing risks ». In : *Biometrics* 71.1, p. 102-113.
- BONFERRONI, Carlo (1935). « Il calcolo delle assicurazioni su gruppi di teste ». In : *Studi in onore del professore salvatore ortu carboni*, p. 13-60.
- BOURSIER, Jérôme et Emmanuel TSOCHATZIS (2021). « Case-finding strategies in non-alcoholic fatty liver disease ». In : *JHEP Reports* 3.2, p. 100219.
- BOURSIER, Jérôme et al. (2008). « Reproducibility of liver stiffness measurement by ultrasonographic elastometry ». In : *Clinical Gastroenterology and Hepatology* 6.11, p. 1263-1269.
- BOURSIER, Jérôme et al. (2016). « Diagnostic accuracy and prognostic significance of blood fibrosis tests and liver stiffness measurement by FibroScan in non-alcoholic fatty liver disease ». In : *Journal of Hepatology* 65.3, p. 570-578.
- BOURSIER, Jérôme et al. (2017). « A stepwise algorithm using an at-a-glance first-line test for the non-invasive diagnosis of advanced liver fibrosis and cirrhosis ». In : *Journal of Hepatology* 66.6, p. 1158-1165.
- BOURSIER, Jérôme et al. (2019). « New sequential combinations of non-invasive fibrosis tests provide an accurate diagnosis of advanced fibrosis in NAFLD ». In : *Journal of Hepatology* 71.2, p. 389-396.

-
- BOX, George et David COX (1964). « An analysis of transformations ». In : *Journal of the Royal Statistical Society : Series B (Methodological)* 26.2, p. 211-243.
- BRETZ, Frank, Torsten HOTHORN et Peter WESTFALL (2016). *Multiple comparisons using R*. Chapman et Hall/CRC.
- BRIER, Glenn et al. (1950). « Verification of forecasts expressed in terms of probability ». In : *Monthly Weather Review* 78.1, p. 1-3.
- BRUDEN, Dana et al. (2017). « Risk of end-stage liver disease, hepatocellular carcinoma, and liver-related death by fibrosis stage in the hepatitis C Alaska Cohort ». In : *Hepatology* 66.1, p. 37-45.
- BYRNE, Christopher et Giovanni TARGHER (2015). « NAFLD : a multisystem disease ». In : *Journal of Hepatology* 62.1, S47-S64.
- CALÈS, Paul et al. (2005). « A novel panel of blood markers to assess the degree of liver fibrosis ». In : *Hepatology* 42.6, p. 1373-1381.
- CARRAT, Fabrice et al. (2019). « Clinical outcomes in patients with chronic hepatitis C after direct-acting antiviral treatment : a prospective cohort study ». In : *The Lancet* 393.10179, p. 1453-1464.
- CASTERA, Laurent, Xavier FORNS et Alfredo ALBERTI (2008). « Non-invasive evaluation of liver fibrosis using transient elastography ». In : *Journal of Hepatology* 48.5, p. 835-847.
- CASTÉRA, Laurent et al. (2005). « Prospective comparison of transient elastography, Fibrotest, APRI, and liver biopsy for the assessment of fibrosis in chronic hepatitis C ». In : *Gastroenterology* 128.2, p. 343-350.
- CHALASANI, Naga et al. (2018). « The diagnosis and management of nonalcoholic fatty liver disease : practice guidance from the American Association for the Study of Liver Diseases ». In : *Hepatology* 67.1, p. 328-357.
- CHAN, Denise et al. (2016). « Sexually acquired hepatitis C virus infection : a review ». In : *International Journal of Infectious Diseases* 49, p. 47-58.
- CLARK, Jeanne, Frederick BRANCATI et Anna Mae DIEHL (2002). « Nonalcoholic fatty liver disease ». In : *Gastroenterology* 122.6, p. 1649-1657.
- COMMENGES, Daniel, Benoit LIQUET et Cécile PROUST-LIMA (2012). « Choice of prognostic estimators in joint models by estimating differences of expected conditional Kullback–Leibler risks ». In : *Biometrics* 68.2, p. 380-387.

-
- COOKE, Graham et al. (2019). « Accelerating the elimination of viral hepatitis : a Lancet Gastroenterology & Hepatology Commission ». In : *The Lancet Gastroenterology & Hepatology* 4.2, p. 135-184.
- COTTER, Thomas et Mary RINELLA (2020). « Nonalcoholic Fatty Liver Disease 2020 : the state of the disease ». In : *Gastroenterology* 158.7, p. 1851-1864.
- COX, David (1972). « Regression models and life-tables ». In : *Journal of the Royal Statistical Society : Series B (Methodological)* 34.2, p. 187-202.
- DAVISON, Beth et al. (2020). « Suboptimal reliability of liver biopsy evaluation has implications for randomized clinical trials ». In : *Journal of Hepatology* 73.6, p. 1322-1332.
- DEMPSTER, Arthur, Nan LAIRD et Donald RUBIN (1977). « Maximum likelihood from incomplete data via the EM algorithm ». In : *Journal of the Royal Statistical Society : Series B (Methodological)* 39.1, p. 1-22.
- DEVAUX, Anthony et al. (2022). « Individual dynamic prediction of clinical endpoint from large dimensional longitudinal biomarker history : a landmark approach ». In : *BMC Medical Research Methodology* 22.1, p. 1-14.
- DUNNETT, Charles et Ajit TAMHANE (1992). « A step-up multiple test procedure ». In : *Journal of the American Statistical Association* 87.417, p. 162-170.
- EASL et al. (2018). « EASL recommendations on treatment of hepatitis C 2018 ». In : *Journal of Hepatology* 69.2, p. 461-511.
- (2021). « EASL Clinical Practice Guidelines on non-invasive tests for evaluation of liver disease severity and prognosis—2021 update ». In : *Journal of Hepatology* 75.3, p. 659-689.
- ESC CARDIOVASC RISK COLLABORATION, SCORE2 WORKING GROUP et al. (2021). « SCORE2 risk prediction algorithms : new models to estimate 10-year risk of cardiovascular disease in Europe ». In : *European Heart Journal* 42.25, p. 2439-2454.
- ESTES, Chris et al. (2018). « Modeling the epidemic of nonalcoholic fatty liver disease demonstrates an exponential increase in burden of disease ». In : *Hepatology* 67.1, p. 123-133.
- FERRER, Loic (2017). « Modélisation et prédiction conjointe de différents risques de progression de cancer à partir des mesures répétées de biomarqueurs ». Thèse de doct.
- FIALLA, Annette, Ove B Schaffalitzky de MUCKADELL et Annmarie TOUBORG LASSEN (2012). « Incidence, etiology and mortality of cirrhosis : a population-based cohort study ». In : *Scandinavian Journal of Gastroenterology* 47.6, p. 702-709.

-
- FINE, Jason et Robert GRAY (1999). « A proportional hazards model for the subdistribution of a competing risk ». In : *Journal of the American Statistical Association* 94.446, p. 496-509.
- FORNS, Xavier et al. (2002). « Identification of chronic hepatitis C patients without hepatic fibrosis by a simple predictive model ». In : *Hepatology* 36.4, p. 986-992.
- GERDS, Thomas, Tianxi CAI et Martin SCHUMACHER (2008). « The performance of risk prediction models ». In : *Biometrical Journal : Journal of Mathematical Methods in Biosciences* 50.4, p. 457-479.
- GOLDBERG, David et al. (2017). « Changes in the prevalence of hepatitis C virus infection, nonalcoholic steatohepatitis, and alcoholic liver disease among patients with cirrhosis or liver failure on the waitlist for liver transplantation ». In : *Gastroenterology* 152.5, p. 1090-1099.
- GRAF, Erika et al. (1999). « Assessment and comparison of prognostic classification schemes for survival data ». In : *Statistics in Medicine* 18.17-18, p. 2529-2545.
- GRAY, Robert (1988). « A class of K-sample tests for comparing the cumulative incidence of a competing risk ». In : *The Annals of Statistics*, p. 1141-1154.
- HAGSTRÖM, Hannes et al. (2020). « Ability of noninvasive scoring systems to identify individuals in the population at risk for severe liver disease ». In : *Gastroenterology* 158.1, p. 200-214.
- HARRELL JR, Frank, Kerry LEE et Daniel MARK (1996). « Multivariable prognostic models : issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors ». In : *Statistics in Medicine* 15.4, p. 361-387.
- HAUSMAN, Jerry et Daniel MCFADDEN (1984). « Specification tests for the multinomial logit model ». In : *Econometrica : Journal of the econometric society*, p. 1219-1240.
- HENDERSON, Charles et al. (1959). « The estimation of environmental and genetic trends from records subject to culling ». In : *Biometrics* 15.2, p. 192-218.
- HENDERSON, Robin, Peter DIGGLE et Angela DOBSON (2000). « Joint modelling of longitudinal measurements and event time data ». In : *Biostatistics* 1.4, p. 465-480.
- HOLM, Sture (1979). « A simple sequentially rejective multiple test procedure ». In : *Scandinavian journal of statistics*, p. 65-70.
- HUANG, Yingxiang et al. (2020). « A tutorial on calibration measurements and calibration models for clinical prediction models ». In : *Journal of the American Medical Informatics Association* 27.4, p. 621-633.

-
- HUNG, Hung et Chin-Tsang CHIANG (2010). « Estimation methods for time-dependent AUC models with survival data ». In : *Canadian Journal of Statistics* 38.1, p. 8-26.
- IMBERT-BISMUT, Françoise et al. (2001). « Biochemical markers of liver fibrosis in patients with hepatitis C virus infection : a prospective study ». In : *The Lancet* 357.9262, p. 1069-1075.
- ISHAK, Kamal (1995). « Histological grading and staging of chronic hepatitis ». In : *Journal of Hepatology* 22, p. 696-699.
- KAPLAN, Edward et Paul MEIER (1958). « Nonparametric estimation from incomplete observations ». In : *Journal of the American Statistical Association* 53.282, p. 457-481.
- KATTAN, Michael et Thomas GERDS (2018). « The index of prediction accuracy : an intuitive measure useful for evaluating risk prediction models ». In : *Diagnostic and Prognostic Research* 2.1, p. 1-7.
- KLEINER, David et al. (2005). « Design and validation of a histological scoring system for nonalcoholic fatty liver disease ». In : *Hepatology* 41.6, p. 1313-1321.
- LASSAILLY, Guillaume et al. (2015). « Bariatric surgery reduces features of nonalcoholic steatohepatitis in morbidly obese patients ». In : *Gastroenterology* 149.2, p. 379-388.
- LINGALA, Shilpa et Marc GHANY (2015). « Natural history of hepatitis C ». In : *Gastroenterology Clinics* 44.4, p. 717-734.
- LUDWIG, Jurgen et al. (1980). « Nonalcoholic steatohepatitis : Mayo Clinic experiences with a hitherto unnamed disease. » In : *Mayo Clinic Proceedings*. T. 55. 7, p. 434-438.
- MANTEL, Nathan (1966). « Evaluation of survival data and two new rank order statistics arising in its consideration ». In : *Cancer Chemother Rep* 50, p. 163-170.
- MARCUS, Ruth, Peritz ERIC et Ruben GABRIEL (1976). « On closed testing procedures with special reference to ordered analysis of variance ». In : *Biometrika* 63.3, p. 655-660.
- MAURER, Willi et Bernhard MELLEIN (1988). « On new multiple tests based on independent p-values and the assessment of their power ». In : *Multiple Hypothesenprüfung/Multiple Hypotheses Testing*. Springer, p. 48-66.
- MCCULLAGH, Peter et John Ashworth NELDER (1989). *Generalized linear models*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. Chapman et Hall/CRC.
- MUTHÉN, Bengt et Kerby SHEDDEN (1999). « Finite mixture modeling with mixture outcomes using the EM algorithm ». In : *Biometrics* 55.2, p. 463-469.

-
- PERAZZO, Hugo et Jean-François DUFOUR (2017). « The therapeutic landscape of non-alcoholic steatohepatitis ». In : *Liver International* 37.5, p. 634-647.
- PETO, Richard et Julian PETO (1972). « Asymptotically efficient rank invariant test procedures ». In : *Journal of the Royal Statistical Society : Series A (General)* 135.2, p. 185-198.
- POYNARD, Thierry et al. (2004). « Prospective analysis of discordant results between biochemical markers and biopsy in patients with chronic hepatitis C ». In : *Clinical Chemistry* 50.8, p. 1344-1355.
- PROUST-LIMA, Cécile et Jeremy TAYLOR (2009). « Development and validation of a dynamic prognostic tool for prostate cancer recurrence using repeated measures of posttreatment PSA : a joint modeling approach ». In : *Biostatistics* 10.3, p. 535-549.
- PROUST-LIMA, Cécile et al. (2014). « Joint latent class models for longitudinal and time-to-event data : a review ». In : *Statistical Methods in Medical Research* 23.1, p. 74-90.
- RATZIU, Vlad et al. (2005). « Sampling variability of liver biopsy in nonalcoholic fatty liver disease ». In : *Gastroenterology* 128.7, p. 1898-1906.
- RIZOPOULOS, Dimitris (2011). « Dynamic predictions and prospective accuracy in joint models for longitudinal and time-to-event data ». In : *Biometrics* 67.3, p. 819-829.
- RIZOPOULOS, Dimitris, Geert VERBEKE et Geert MOLENBERGHS (2008). « Shared parameter models under random effects misspecification ». In : *Biometrika* 95.1, p. 63-74.
- ROBERTS, Eve et Latifa YEUNG (2002). « Maternal-infant transmission of hepatitis C virus infection ». In : *Hepatology* 36.S1, S106-S113.
- ROMERO-GÓMEZ, Manuel, Shira ZELBER-SAGI et Michael TRENELL (2017). « Treatment of NAFLD with diet, physical activity and exercise ». In : *Journal of Hepatology* 67.4, p. 829-846.
- ROSENBERG, William et al. (2004). « Serum markers detect the presence of liver fibrosis : a cohort study ». In : *Gastroenterology* 127.6, p. 1704-1713.
- SCHOENFELD, David (1982). « Partial residuals for the proportional hazards regression model ». In : *Biometrika* 69.1, p. 239-241.
- SCHOOP, Rotraut et al. (2011). « Quantifying the predictive accuracy of time-to-event models in the presence of competing risks ». In : *Biometrical Journal* 53.1, p. 88-112.
- SCHWARZ, Gideon et al. (1978). « Estimating the dimension of a model ». In : *Annals of Statistics* 6.2, p. 461-464.

-
- SCHWENZER, Nina et al. (2009). « Non-invasive assessment and quantification of liver steatosis by ultrasound, computed tomography and magnetic resonance ». In : *Journal of Hepatology* 51.3, p. 433-445.
- SHAFFER, Juliet Popper (1986). « Modified sequentially rejective multiple test procedures ». In : *Journal of the American Statistical Association* 81.395, p. 826-831.
- SJÖSTRÖM, Lars et al. (2004). « Lifestyle, diabetes, and cardiovascular risk factors 10 years after bariatric surgery ». In : *New England Journal of Medicine* 351.26, p. 2683-2693.
- SPEARMAN, Wendy et al. (2019). « Hepatitis C ». In : *The Lancet* 394.10207, p. 1451-1466.
- SRIVASTAVA, Ankur et al. (2019a). « Cost-comparison analysis of FIB-4, ELF and fibroscan in community pathways for non-alcoholic fatty liver disease ». In : *BMC Gastroenterology* 19.1, p. 1-15.
- SRIVASTAVA, Ankur et al. (2019b). « Prospective evaluation of a primary care referral pathway for patients with non-alcoholic fatty liver disease ». In : *Journal of Hepatology* 71.2, p. 371-378.
- STERLING, Richard et al. (2006). « Development of a simple noninvasive index to predict significant fibrosis in patients with HIV/HCV coinfection ». In : *Hepatology* 43.6, p. 1317-1325.
- THE FRENCH METAVIR COOPERATIVE STUDY GROUP (1994). « Intraobserver and interobserver variations in liver biopsy interpretation in patients with chronic hepatitis C ». In : *Hepatology* 20.1, p. 15-20.
- TRINCHET, Jean-Claude et al. (2015). « Complications and competing risks of death in compensated viral cirrhosis (ANRS CO12 CirVir prospective cohort) ». In : *Hepatology* 62.3, p. 737-750.
- TSIATIS, Anastasios et Marie DAVIDIAN (2004). « Joint modeling of longitudinal and time-to-event data : an overview ». In : *Statistica Sinica*, p. 809-834.
- UNO, Hajime et al. (2007). « Evaluating prediction rules for t-year survivors with censored regression models ». In : *Journal of the American Statistical Association* 102.478, p. 527-537.
- VAN HOUWELINGEN, Hans (2007). « Dynamic prediction by landmarking in event history analysis ». In : *Scandinavian Journal of Statistics* 34.1, p. 70-85.
- VAN HOUWELINGEN, Hans et Hein PUTTER (2011). *Dynamic prediction in clinical survival analysis*. CRC Press.

-
- VERBEKE, Geert et Emmanuel LESAFFRE (1996). « A linear mixed-effects model with heterogeneity in the random-effects population ». In : *Journal of the American Statistical Association* 91.433, p. 217-221.
- VERBEKE, Geert et Geert MOLENBERGHS (2009). *Linear mixed models for longitudinal data*. Springer Science & Business Media.
- WAI, Chun-Tao et al. (2003). « A simple noninvasive index can predict both significant fibrosis and cirrhosis in patients with chronic hepatitis C ». In : *Hepatology* 38.2, p. 518-526.
- WESTBROOK, Rachel et Geoffrey DUSHEIKO (2014). « Natural history of hepatitis C ». In : *Journal of hepatology* 61.1, S58-S68.
- WESTFALL, Peter et al. (1999). *Multiple Comparisons and Multiple Tests Using SAS*. SAS Institute.
- WHO (2017). *World Health Organization : Global hepatitis report, 2017*. URL : <https://www.who.int/publications/i/item/9789241565455> (visité le 14/04/2022).
- (2021). *World Health Organization : Hepatitis C*. URL : <https://www.who.int/news-room/fact-sheets/detail/hepatitis-c> (visité le 14/04/2022).
- WONG, Grace Lai-Hung (2018). « Non-invasive assessments for liver fibrosis : The crystal ball we long for ». In : *Journal of Gastroenterology and Hepatology* 33.5, p. 1009-1015.
- YOUNOSSI, Zobair (2019). « Non-alcoholic fatty liver disease—a global public health perspective ». In : *Journal of Hepatology* 70.3, p. 531-544.
- YOUNOSSI, Zobair et al. (2011). « Changes in the prevalence of the most common causes of chronic liver diseases in the United States from 1988 to 2008 ». In : *Clinical Gastroenterology and Hepatology* 9.6, p. 524-530.
- YOUNOSSI, Zobair et al. (2019). « Global perspectives on nonalcoholic fatty liver disease and nonalcoholic steatohepatitis ». In : *Hepatology* 69.6, p. 2672-2682.
- YPMA, Tjalling (1995). « Historical development of the Newton–Raphson method ». In : *SIAM review* 37.4, p. 531-551.

VALORISATION SCIENTIFIQUE ET ENSEIGNEMENTS

Publications

1. Jérôme Boursier, Hannes Hagström, Mattias Ekstedt, Clémence Moreau, Martin Bonacci, Sandrine Cure, Javier Ampuero, Patrik Nasr, Lilian Tallab, Clémence M. Canivet, Stergios Kechagias, Yolanda Sánchez, Eloise Dincuff, Ana Lucena, Marine Roux, Jérémie Riou, Aldo Trylesinski, Manuel Romero-Gomez. « Non-invasive tests accurately stratify patients with NAFLD based on their risk of liver-related events ». Publié dans : *Journal of Hepatology* (01/2022).

Contribution personnelle : data management, analyse statistique.

2. Clémence Moreau, Marine Roux, Jérémie Riou, Pierre Nahon, Fabrice Carrat, Jérôme Boursier, the ANRS CO12 CirVir group and the ANRS CO22 HEPATHER group. « Dynamic Personalized Prediction of the Individual Liver-Related Risk after Sustained Viral Response in HCV patients ». En cours de soumission (2022).

Contribution personnelle : data management, analyse statistique, rédaction du premier draft de l'article, création de l'application Rshiny.

3. Clémence Moreau, Jérémie Riou, Marine Roux. « Comparison of multiple performance criteria for dynamics predictions ». En cours de soumission (2022).

Contribution personnelle : développement méthodologique, études de simulation, application, rédaction du premier draft de l'article, programmation R de la procédure disponible sur GitHub.

Communications scientifiques orales

1. Clémence Moreau, Jérémie Riou, Marine Roux (Juin 2021). « Comparaison de multiples critères de performance de prédictions dynamiques ». 52^{èmes} Journées de Statistique de la Société Française de Statistique, Visioconférence.

-
2. Clémence Moreau, Jérémie Riou, Marine Roux (Juillet 2021). « Comparison of multiple dynamic predictive accuracies ». *42^{ème} Conférence de la Société Internationale de Biostatistique Clinique*, Visioconférence.

Communications scientifiques affichées

1. Clémence Moreau, Jérôme Boursier, Jérémie Riou (Octobre 2019). « Impact de la progression de la fibrose hépatique chez les patients atteints de NAFLD ». *Journées du GDR « Statistiques et Santé »*, Paris (France).
2. Clémence Moreau, Jérôme Boursier, Jérémie Riou, Marine Roux (Décembre 2019). « Impact de la progression de la fibrose hépatique chez les patients atteints de NAFLD ». *Journée de recherche biomédicale Angers-Tours*, Tours (France).
3. Clémence Moreau, Jérôme Boursier, Jérémie Riou, Marine Roux (Septembre 2020). « Impact de la progression de la fibrose hépatique chez les patients atteints de NAFLD ». *14^{ème} Conférence Francophone d'Épidémiologie Clinique*, Visioconférence.
4. Jérôme Boursier, Hannes Hagström, Mattias Ekstedt, Clémence Moreau, Martin Bonacci, Sandrine Cure, Aldo Trylesinski, Manuel Romero-Gómez (Juin 2021). « FIB-4 and FibroScan can identify patients with Non Alcoholic Fatty Liver Disease who are at risk of liver-related events : Results of a longitudinal analysis with comparison with liver biopsy ». *International Liver Congress 2021*, Visioconférence.
5. Clémence Moreau, Jérôme Boursier, Jérémie Riou, Marine Roux (Décembre 2021). « Prédiction de risque personnalisée après une réponse virale soutenue chez les patients cirrhotiques infectés par le virus de l'hépatite C ». *Journée de recherche biomédicale Angers-Tours*, Angers (France).
6. Clémence Moreau, Jérémie Riou, Marine Roux (Décembre 2021). « Comparison of multiple dynamic predictive accuracies ». *4^{èmes} Journées Scientifiques de l'École doctorale Biologie Santé Bretagne Loire*, Brest (France).
7. Clémence Moreau, Jérémie Riou, Marine Roux (Juin 2022). « Comparison of multiple dynamic predictive accuracies ». *Journée SFR ICAT 2022 : Maladies Dysmétaboliques*, Angers (France).

Enseignement - Encadrement

- * Contrat Doctorante Contractuelle à Activité Complémentaire d'Enseignement réalisé au sein de la faculté de Mathématiques de l'Université d'Angers (2020-2021). Intervention type conférence professionnelle CM/TP M2 Data Science (6h), TD « Probabilités » L2 SVT (16h), CM/TD/Examen « Mathématiques : notions de probabilités et statistiques » L1 SVT option Cursus Master Ingénierie (10h).
- * Encadrement d'un stage de 3 mois d'une étudiante de l'ENSAI pour travailler sur la programmation R de l'AUC dynamique (06-08/2020).

GitHub

- * L'implémentation R de la procédure de comparaison de plusieurs critères de performance dynamiques développée lors du Chapitre 5 est disponible sur https://github.com/ClemenceMoreau/Comparison_multiple_performance_criteria.

Activités complémentaires

- * Analyse et rapport statistiques réalisés pour le laboratoire pharmaceutique *Intercept Pharma Europe Ltd.* (Mai-Juin 2020).
- * Participation à la 8^{ème} *Conférence du Channel Network* (Avril 2021). Participation au *short course* « Co-data learning in high dimensional prediction problems » animé par Mirrelijn van Nee, Magnus Münch et Mark van de Wiel. Visioconférence.
- * Participation au *Workshop Inserm* « Recent advances in statistical analysis of survival data » (Mai 2021). Visioconférence.
- * Participation au *short course* « Joint Models under the Bayesian approach » animé par Dimitris Rizopoulos lors de la 42^{ème} *Conférence de la Société Internationale de Biostatistique Clinique* (Juillet 2021). Visioconférence.

Titre : Modélisation et prédiction dynamiques du risque de complications dans le cadre de pathologies hépatiques

Mot clés : AUC & Brier Score dynamiques ; Hépatite C ; Modélisation & prédiction dynamiques ; NAFLD ; Tests non-invasifs.

Résumé : La prise en charge des patients présentant une fibrose du foie repose sur l'évaluation histologique de ce dernier. Pour cela, l'examen de référence est la biopsie, mais étant donné les risques, les biais d'échantillonnage et d'observateur qu'elle comporte, les tests non-invasifs lui sont aujourd'hui préférés. Ils connaissent cependant des performances diagnostiques limitées. Un moyen de dépasser cette limite serait de ne plus prendre en charge les patients selon l'état lésionnel de leur foie mais plutôt selon leur pronostic. Cela permettrait une prise en charge personnalisée et totalement non-invasive pour ces malades. L'objectif de cette thèse est d'apporter des réponses statistiques à ce changement de paradigme clinique. Ainsi, un premier travail consiste à comparer les performances pronostiques de la

biopsie face aux tests non-invasifs dans le cadre de la NAFLD. Le second projet de cette thèse propose de prédire les complications hépatiques au sein de l'hépatite C chronique à l'aide d'un modèle conjoint à classes latentes. Cette modélisation dite dynamique permet d'obtenir des prédictions de risque personnalisées pour les patients, mises à jour tout au long de leur suivi. Enfin, un troisième travail introduit un nouvel outil statistique permettant la comparaison de plusieurs critères de performance dynamiques afin de déterminer le meilleur biomarqueur pronostique à suivre dans le temps ou le meilleur modèle dynamique. Cette nouvelle méthode a été appliquée chez des patients atteints de NAFLD afin de comparer les capacités pronostiques de plusieurs tests non-invasifs pour prédire des complications hépatiques.

Title: Dynamic modeling and prediction of the complication risk in liver pathologies

Keywords: Dynamic AUC & Brier Score; Dynamic modeling & prediction; Hepatitis C; NAFLD; Non-invasive tests.

Abstract: The follow-up of patients with liver fibrosis is based on histological evaluation of the liver. The gold standard is biopsy, but due to risk, sampling and observer bias, non-invasive tests are now preferred. However, their diagnostic performance is limited. It would be possible to overcome this limitation by no longer monitoring patients according to the lesion status of their liver but rather according to their prognosis. This would allow a personalized and totally non-invasive follow-up of patients. The objective of this thesis is to provide statistical answers to this change in clinical paradigm. Thus, a first work consists in comparing the prognostic performances of biopsy versus non-

invasive tests in the context of NAFLD. The second project of this thesis proposes to predict liver complications in chronic hepatitis C using a joint latent class model. This so-called dynamic modelling provides personalized risk predictions for patients, updated throughout their follow-up. Finally, a third work introduces a new statistical tool allowing the comparison of several dynamic performance criteria in order to determine the best prognostic biomarker to follow over time or the best dynamic model. This new method was applied in patients with NAFLD to compare the prognostic capabilities of several non-invasive tests for predicting liver complications.

