



**HAL**  
open science

## **Liver fibrosis staging by computed tomography: prospective randomized multicentric evaluation of image analyses**

Carole Vitellius, Anita Paisant, Adrien Lannes, Julien Chaigneau, Frédéric Oberti, Jérôme Lebigot, Isabelle Fouchard, Jérôme Boursier, Pascal David, Christophe Aubé, et al.

### ► To cite this version:

Carole Vitellius, Anita Paisant, Adrien Lannes, Julien Chaigneau, Frédéric Oberti, et al.. Liver fibrosis staging by computed tomography: prospective randomized multicentric evaluation of image analyses. Clinics and Research in Hepatology and Gastroenterology, 2021, pp.101797. 10.1016/j.clinre.2021.101797 . hal-03417068

**HAL Id: hal-03417068**

**<https://univ-angers.hal.science/hal-03417068v1>**

Submitted on 5 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.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

## TITLE PAGE

# Liver fibrosis staging by computed tomography: prospective randomized multicentric evaluation of image analyses

**Short Title:** Liver fibrosis staging by CT

Carole Vitellius<sup>1,2</sup>, Anita Paisant<sup>2,3</sup>, Adrien Lannes<sup>1,2</sup>, Julien Chaigneau<sup>2</sup>, Frédéric Oberti<sup>1,2</sup>, Jérôme Lebigot<sup>2,3</sup>, Isabelle Fouchard<sup>1,2</sup>, Jérôme Boursier<sup>1,2</sup>, Pascal David<sup>4</sup>, Christophe Aubé<sup>2,3</sup>, Paul Calès<sup>1,2</sup> and the CDMIR group\*

### Authors' institutions:

- (1) Hepato-gastroenterology department, Angers University Hospital, Angers, France
- (2) HIFIH Laboratory UPRES EA3859, SFR 4208, Angers University, Angers, France
- (3) Radiology department, Angers University Hospital, Angers, France
- (4) Radiology centre, 24 Couscher street, 49400 Saumur, France

\* The list of contributions is provided at the end of the manuscript.

**Financial Support:** Institutional. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

### Abbreviations

CDMIR: comprehensive dynamic medical imaging report

CLD: chronic liver disease

CT: computed tomography

F: Metavir liver fibrosis stage

FR: radiological liver fibrosis stage

HGE: hepato-gastroenterology

ICCC: intra-class correlation coefficient

NAFLD: non-alcoholic fatty liver disease

NIT: non-invasive test

NPV: negative predictive value

PPV: positive predictive value

**Correspondence:**

Paul Calès, Hepatology, CHU, 49933 Angers Cedex 09, France. Tel: (33) 2 41 35 34 10, Fax:

(33) 2 41 35 41 19, E-mail: paul.cales@univ-angers.fr

<http://orcid.org/0000-0003-4866-5274>

**Authors' contributions:**

Carole Vitellius: data collection, critical revision of the manuscript

Anita Paisant: data collection, critical revision of the manuscript

Adrien Lannes: data collection, critical revision of the manuscript

Julien Chaigneau: data collection (morphometric analysis)

Frédéric Oberti: data collection

Jérôme Lebigot: data interpretation

Isabelle Fouchard: data collection

Jérôme Boursier: data collection

Pascal David: data interpretation

Christophe Aubé: study supervision, data interpretation

Paul Calès: study supervision, data collection, analysis and interpretation of data, drafting of the manuscript, statistics, corresponding author

CDMIR group: data interpretation

**Item count:**

Main text excluding references: 4166 words; 4 tables, 3 figures; number of references: 31.

Title: 112 characters, short title: 28 characters (<30 spaces including spaces)

Supplemental Material including 18 tables and 6 figures

**Declaration of Conflict of Interest**

None related with the topic for all authors.

**Authors' e-mails:**

Carole.Vitellius@chu-angers.fr

Anita.Paisant@chu-angers.fr

Adrien.Lannes@chu-angers.fr

julien.chaigneau@univ-angers.fr

FrOberti@chu-angers.fr

JeLebigot@chu-angers.fr

Isabelle.Fouchard@chu-angers.fr

JeBoursier@chu-angers.fr

p.david3@wanadoo.fr

ChAube@chu-angers.fr

**Writing Assistance:** Kevin L. Erwin for English proofreading (institutional support)

## **ABSTRACT**

**Aim:** Liver fibrosis staging is essential. We prospectively evaluated the liver fibrosis staging performance of computed tomography (CT).

**Methods:** 70 hepato-gastroenterology clinicians were randomized into three stratified groups with different image analyses of radiological semiology, i.e., on raw images (group 1) and on expert-annotated (group 2) and computerized-morphometry-enriched (group 3) images.

Radiological fibrosis staging based on seven simple descriptors into four stages equivalent to Metavir stages (F0/1, F2, F3, F4=cirrhosis) was determined at baseline and after image analyses in 10 patients with chronic liver diseases (two per F) concordant for four independent fibrosis stagings including Metavir. 23,800 CT images were analysed, providing 1400 fibrosis stagings.

**Results:** Fibrosis staging: overall (3 groups) accuracy (correct classification rate) was, baseline: 43%, post-analysis: 60% ( $p < 0.001$ ) without significant progression in group 1 (6%,  $p = 0.207$ ) contrary to groups 2 (34%,  $p < 0.001$ ) and 3 (13%,  $p = 0.007$ ). Cirrhosis diagnosis: overall accuracy was, baseline: 84%, post-analysis: 89% ( $p < 0.001$ ) without significant progression in group 1 (0%,  $p = 1$ ) contrary to groups 2 (8%,  $p = 0.009$ ) and 3 (7%,  $p = 0.015$ ). Baseline AUROCs were good ( $\geq 0.83$ ) for marked fibrosis ( $F \geq 3$  or cirrhosis) in all groups. Post-analysis AUROCs became excellent ( $\geq 0.89$ ) in group 2 for all diagnostic targets ( $\geq 0.98$  for  $F \geq 3$  and cirrhosis) and in group 3 for cirrhosis. In post-analysis group 2, discrimination between all F was excellent (especially, F1 from F0) with an Obuchowski index at 0.87. Negative and positive predictive values for marked fibrosis were 98% and 95%, respectively.

**Conclusion:** Simple CT descriptors accurately discriminate all Metavir liver fibrosis stages.  
(n=250 words)

**Keywords:** computed tomography, liver fibrosis, cirrhosis, staging, non-invasive diagnosis

## INTRODUCTION

The evaluation of liver fibrosis is essential in the management of patients with chronic liver disease (CLD). However, due to multiple biases, liver biopsy cannot be considered a gold standard, but rather only a best standard [1]. Many teams have developed non-invasive assessments of liver fibrosis, particularly blood fibrosis tests and transient elastography [2]. Radiological imaging could provide a double advantage of enabling both a fibrosis assessment based on dysmorphia and an exploration of the entire liver in a single examination [3]. Computed tomography (CT) is a widely available and frequently employed imaging technology, but its potential in this setting has received little attention and, despite such recent techniques as texture analysis [4], no classification of liver fibrosis stages is available for it. Our objective was to evaluate the contribution of descriptors for fibrosis staging while avoiding classical biases: observer expertise and liver biopsy as reference. Our ensuing working hypothesis was two-fold. First, a way to avoid expertise bias would be to employ a naive population of investigators. Second, the bias due to liver biopsy might be avoided by using Metavir F concordant over several independent non-invasive tests (NITs) as a reference, as done in diagnostic studies without gold standards [5].

Thus, our primary objective with the present proof of concept study was to validate radiological staging for liver fibrosis. Our secondary objectives were to compare image analysis fibrosis stagings done by radiologist to those done by computer, and stagings done by clinicians to those done by radiologists. The present study, called comprehensive dynamic medical imaging report (CDMIR), was a prospective diagnostic study with stratified randomization of observers and patients.

## **METHODS**

### **Overall presentation**

Images from 10 patients were compared under three formats: native raw images as internal controls; images enriched by expert annotations; and images enriched by computerized morphometry. The main evaluators were naive hepato-gastroenterology (HGE) residents.

### **Patients**

Patients were selected from the MALAH2 CLD population (details in Supplemental Material). A patient flowchart is provided in [Figure S1](#) in Supplemental Material. In a preliminary study, liver CT images were evaluated by three senior radiologists in 106 patients [6]. Ten of those 106 patients were included in the present CDMIR study, based on fibrosis staging concordance over four independent diagnostic tools for liver fibrosis (Metavir F, blood test, elastometry and CT morphometry, [Table S1](#) in Supplemental Material) to optimize the histological Metavir reference for liver fibrosis [1, 5]. Patients were stratified for the five Metavir F stages measured on liver specimens, i.e., two in each Metavir F stage. The main characteristics of the 10 patients included in the present study were: age:  $56.8 \pm 15.7$  years, male sex: 40.0%, body mass index:  $30.4 \pm 5.5$  kg/m<sup>2</sup>, non-alcoholic fatty liver disease (NAFLD): 80.0%. Their individual liver fibrosis descriptors are provided in [Tables S1](#) and [S2](#).

### **Investigators**

*Observers* - Seventy HGE residents (“observers” hereafter) working in five university hospital centres (“centre” hereafter) participated in the study. Their inclusion into three groups was randomly stratified according to two criteria: centre (n=5) and internship year (n=4). There were no significant differences in observer characteristics between the three groups (Table S3).

*Radiologists* - Four senior radiologists participated in the study. First, an HGE expert radiologist (9 years of experience; “descriptor expert” hereafter) selected and annotated the images used for observer evaluation. Three radiologists evaluated images: two specialists in HGE (22 and 28 years of experience) and one non-specialist in HGE (29 years in general radiology). These three radiologists (“evaluators” hereafter) independently evaluated the same raw (i.e. those without annotations) CT images as the HGE observers did. Unlike HGE observers, the radiologists had no time constraints.

## **Liver imaging**

*Classical semiology* - Based on findings reported by expert HGE radiologists in previous studies [7-10], seven conventional, simple descriptors of liver fibrosis and cirrhosis were selected [6]: 1/ liver edge (normal, moderate or marked irregularity), 2/ right posterior notch (present/absent) [8], 3/ segment I hypertrophy (absent, moderate or marked), 4/ segment IV hypotrophy (absent, moderate or marked), 5/ splenomegaly (absent or present according to spleen size measurement along its major axis), 6/ coarse appearance of liver parenchyma (absent, moderate or marked) and 7/ portosystemic collaterals (absent, minimal or marked) (Figure 1A).

*CT* - CT images without or with contrast medium injection at portal time were used. The technique is detailed in the Supplemental Material.



## **New CT classification of liver fibrosis**

The CT classification of liver fibrosis stages was developed during a consensus meeting between the most experienced radiologist and the principal investigator. Histological Metavir F staging was the reference [11] and the seven previous radiological descriptors were considered. The extreme radiological liver fibrosis stages ( $F_R$ ) were defined first as normal ( $F_{R0}$ ) and cirrhosis ( $F_{R4}$ ) according to the absence or presence of classical signs. Thereafter, two intermediate stages were defined, i.e., moderate or marked abnormalities without a firm cirrhosis diagnosis ( $F_{R3}$ ) and moderate abnormalities with no doubt as to the absence of cirrhosis ( $F_{R2}$ ). A theoretical  $F_{R1}$  stage was grouped with  $F_{R0}$  as the experts estimated that their distinction was not possible in principle. Thus, the four  $F_R$  stages were: stage  $F_{R0/1}$  for a (near) normal liver,  $F_{R2}$  for significant fibrosis,  $F_{R3}$  for marked fibrosis or possible cirrhosis and  $F_{R4}$  for definitive cirrhosis.

*Morphometric analysis* – This technique included computerized image analysis as described elsewhere [12]. From the preliminary study, we selected images of six native descriptors among 43 morphometric descriptors [6]. Coloured masks were possible ([Figure 1B](#)). The six descriptors were the independent predictors of fibrosis stages or cirrhosis included by multivariate analysis into two diagnostic scores (for Metavir F or cirrhosis). The ensuing eight measurements (the six descriptors plus the two scores) are detailed in [Table S2](#).

## **Study design**

### *Study conduct*

We conducted the observer evaluation during a workshop session performed in Angers on 16 October 2018. All participants were blinded to the biological/clinical background of patients until the end of the evaluation session.

### ***Definitions***

*Objectives* - The primary objective was to develop radiological staging for liver fibrosis and evaluate its performance (mainly discrimination). The secondary objectives were to a) compare diagnostic accuracies for fibrosis staging between images enriched by radiologist annotations and images enriched by morphometric analysis, and between HGE observers and radiologists; b) measure fibrosis staging agreement between observers and liver pathology (i.e. calibration) or observers themselves (i.e. reproducibility); c) compare agreement of fibrosis stagings obtained by the observers on CT and those obtained by different pathologists on liver biopsy.

*Outcome measurements* - The first primary outcome measurement was the rate of correctly-classified patients (accuracy in percentages hereafter). This was expressed with four diagnostic descriptors: pre- and post-analysis accuracies and the accuracy course expressed either as a qualitative descriptor for observers (stable, regression, progression) or as a quantitative descriptor, i.e., post-analysis accuracy - baseline accuracy, called *net change* hereafter. The classification reference was Metavir F stages by liver biopsy. The main diagnostic targets were fibrosis staging and cirrhosis. The second primary outcome measurement was the  $F_R$  discrimination between Metavir F either for binary fibrosis targets ( $\geq F1$ : fibrosis,  $\geq F2$ : significant fibrosis,  $\geq F3$ : marked fibrosis, F4: cirrhosis) by AUROC or all F by Obuchowski index or between two adjacent F.

The secondary outcome measurements were observer agreement, either interobserver (reproducibility) or with Metavir F staging (i.e. calibration of radiological staging), the

comparison of accuracies between HGE observers and radiologists, and the comparison of agreement between observers and pathologists. This last outcome is reported in the Supplementary Material.

*Study* - This work was a prospective (data collection organized upstream of the performance of index tests and the reference standard with prospective aim) [13] multicentre study with double stratified randomization (patients and observers). It was a type 1a predictive study according to TRIPOD classification [14]. Data were reported according to STARD [13] and Liver FibroSTARD [15] statements.

### ***Evaluation session***

The design of the observer evaluation session is summarized in [Table 1](#). Details are reported in Supplemental Material with examples in [Figure 1](#). Briefly, the observers were randomized into three groups according to the images evaluated. Images were evaluated twice. First, the same raw images in all groups. Second, the same images with different formats according to groups: control group 1C with raw images, group 2A with annotated images, and group 3B with images analysed by bio-morphometry.

### **Regulatory aspects**

*Patients* - The CDMIR study was conducted in the setting of the SNIFF cohort, which included all patients with CLD managed in our centre. Written, informed consent was obtained from all patients.

*Observers* - the study, termed an ‘educational research study’, was not considered as ‘Research Involving the Human Person’ as defined in article R1121-1 of the French Public Health Code, and thus no Ethics Committee approval was required for the observers.

## **Statistical analysis**

Quantitative variables were expressed as mean  $\pm$  standard deviation. Comparisons between groups were done using an ANOVA parametric test or Chi<sup>2</sup> test. The 2 to 2 comparisons were carried out using an unpaired t-test for continuous variables and a Fischer test for qualitative variables. For intra-group comparisons (baseline and post-analysis data), a paired t-test was used. AUROCs were compared by the paired Delong test or unpaired reduced gap test. Discrimination between the five Metavir F stages was described by the Obuchowski index [16] (details in Supplemental Material). The comparisons were also performed using non-parametric tests. These latter provided similar results and are thus not reported here. No weighting was introduced for multiple comparisons [17]. Multivariate analyses were performed by backward binary logistic regression with or without interaction terms, excluding colinearity >0.8. Inter-observer agreement was assessed by the kappa index and the intra-class correlation coefficient (ICCC). The sample size is discussed in the Supplemental Material. Statistical analyses were performed mainly with SPSS software version 18.0 (IBM, Armonk, NY, USA).

## RESULTS

### Accuracy

#### *Fibrosis staging*

*Individual observer description* - Individual baseline and post-analysis accuracies are detailed in [Figure S2](#). In group 1C (n=22, with the same images in the two steps), mean accuracy (in 10 patients) was stable in 4 observers, progressed in 11 and regressed in 7. In group 2A (n=21), mean accuracy progressed in all observers. In group 3B (n=27), mean accuracy was stable in 7 observers, progressed in 15 and regressed in 5. So, the main trend was 100% progression in group 2A vs around 50% in other groups ( $p < 0.001$ , [Table S4](#)).

*Univariate analysis* ([Table 2A](#) and [Figure 2A](#)) - The overall (all groups merged) baseline  $F_R$  accuracy for Metavir fibrosis stages was 43% without significant differences between the three groups ( $p = 0.186$ ). The overall post-analysis accuracy was 60% with significant differences between the three groups ( $p < 0.001$ ). Specifically, post-analysis accuracy was higher in group 2A (81%) compared to group 1C (51%,  $p < 0.001$ ) and to group 3B (52%,  $p < 0.001$ ), but it did not significantly differ between groups 1C and 3B ( $p = 0.786$ ). Therefore, the net change (absolute difference) was significantly higher in group 2A than in the others.

*Multivariate analysis* - Metavir F had a positive influence ( $p < 0.001$ ) and centre 5 had a negative influence ( $p = 0.017$ ) on baseline accuracy for fibrosis  $F_R$  staging but not the group ([Table S5A](#)). For post-analysis accuracy, group 2A ( $p < 0.001$ ) had a positive influence in addition to the two previous predictors.

#### *Cirrhosis diagnosis*

*Univariate analysis* (Table 2B and Figure 2A) - The overall baseline  $F_R$  accuracy for cirrhosis was 84% without significant differences between the three groups ( $p=0.163$ ). The overall post-analysis accuracy was 89% with significant differences between the three groups ( $p<0.001$ ). Specifically, the post-analysis accuracy was higher in group 2A (96%) compared to group 1C (82%,  $p<0.001$ ) or 3B (90%,  $p=0.012$ ) and it did differ significantly between groups 1C and 3B ( $p=0.012$ ). Therefore, the net change was significantly higher in groups 2A and 3B than it was in controls.

*Multivariate analysis* - Metavir F ( $p<0.001$ ) and centre 3 ( $p=0.023$ ) had a negative influence on baseline accuracy for cirrhosis but not the group (Table S5B). For post-analysis accuracy, Metavir F ( $p<0.001$ ) had a negative influence while groups 2A ( $p<0.001$ ) and 3B ( $p=0.008$ ) had positive influences.

### ***Fibrosis staging vs cirrhosis diagnosis***

As expected, the  $F_R$  accuracy was significantly higher for cirrhosis than for fibrosis staging in the baseline ( $p<0.001$ ) and post-analysis ( $p<0.001$ ) steps (Table 2C). In contrast, the net change was significantly lower for cirrhosis than it was for fibrosis staging ( $p<0.001$ ).

### ***Sensitivity analysis***

Centre

Accuracies as a function of centres and groups are shown in Figures 2B-D.

*Fibrosis staging* - There was no significant difference in baseline accuracy between centres ( $p=0.258$ ) (Table S6A). There was a significant difference in post-analysis accuracy

( $p < 0.001$ ) or net change ( $p = 0.025$ ) between centres with a significantly lower progression in centre 5 vs other centres due to a negative net change in group 3B (Table S7).

*Cirrhosis diagnosis* - There was no significant difference in baseline accuracy ( $p = 0.134$ ) (Table S6B), post-analysis accuracy ( $p = 0.655$ ), and net change ( $p = 0.554$ ) between the five centres.

#### Internship year

Regarding the four accuracy descriptors, there were no significant differences for fibrosis staging and cirrhosis diagnosis according to the internship years (Table S8).

#### Metavir F

*Fibrosis staging* - Baseline and post-analysis accuracies for fibrosis staging were significantly different as a function of Metavir F ( $p < 0.001$ ) (Table S9). Both accuracies were the highest in Metavir F4 and the lowest in Metavir F1 whatever the group (Table S10). The net change was significant, with progression in all Metavir F (borderline significance in F4).

*Cirrhosis diagnosis* - Baseline and post-analysis accuracies for cirrhosis diagnosis were significantly different as a function of Metavir F with the highest in Metavir  $F \leq 2$  and the lowest in Metavir  $F \geq 3$  (Table S9). Therefore, the accuracies for cirrhosis were significantly higher in patients without than in patients with cirrhosis in all groups (Table S11). However, there was a notable exception in group 2A: post-analysis accuracy was 96% in  $F \leq 3$  vs 98% in F4 ( $p = 1$ ). The net change was significant (progression) only in F3 (Table S9).

## AUROC

The AUROCs of observer fibrosis stagings were measured for four diagnostic binary targets according to Metavir F cut-offs ([Table 3](#)):  $F \geq 1$  (see discussion for this unexpected target),  $F \geq 2$ ,  $F \geq 3$  and  $F = 4$ . In baseline evaluation, AUROCs were not significantly different between groups (except a borderline significant decrease in group 3B for  $F \geq 1$ ). The post-analysis AUROCs of group 2A were significantly increased vs the baseline AUROCs of that group and vs the post-analysis AUROCs of the other groups. The post-analysis AUROC for cirrhosis of group 3B was significantly higher than the post-analysis AUROCs of group 1C. Finally, baseline overall (all groups merged) AUROCs were satisfactory ( $\geq 0.83$ ) for marked fibrosis ( $F \geq 3$  or cirrhosis) whereas post-analysis AUROCS were excellent ( $\geq 0.89$ ) in group 2A for all diagnostic targets (especially  $\geq 0.98$  for  $F \geq 3$  and cirrhosis) and in group 3B for cirrhosis (0.93). The Obuchowski index, reflecting discrimination between Metavir F stages, significantly ( $p < 0.001$ ) progressed in groups 2A and 3B with the greater progression in group 2A, from 0.66 to 0.87 ([Table 3](#)). This provided excellent negative (NPV: 98%) and positive (PPV: 95%) predictive values for marked fibrosis ( $F \geq 3$ ) in post-analysis group 2A ([Table S12](#)).

## Observer agreement

*Inter-observer agreement* - Inter-observer agreement for fibrosis  $F_R$  staging was assessed by ICC in the baseline and post-analysis evaluations or both steps merged in each group ([Table S13](#)). In the baseline evaluation, agreement was fair in groups 1C and 3C and slightly higher in group 2A. Agreement was not significantly different between evaluation steps in group 1C, as expected. Agreement significantly improved between the two steps in groups 2A and 3B,



especially in the former, where it became excellent, reflecting the accuracy results. Indeed, accuracy and ICC were highly correlated ( $r_p=0.927$ ,  $p=0.008$ ) (Figure S3).

*Agreement between observer  $F_R$  and Metavir F* - Overall agreement (i.e. calibration of radiological staging) was significantly improved in the post-analysis step vs the baseline step in groups 2A and 3B (Table S14). Figure 3 shows that calibration was satisfactory for all Metavir F stages only in the post-analysis step of group 2A, with  $r_s=0.914$  and  $ICC=0.951$ . Of note, calibration improved for all F in post-analysis group 2A (Table S15). Thus, in group 2A, post-analysis calibration became excellent in Metavir F2 to F4 while calibration was good in Metavir F1 and fair in Metavir F0. In addition, the discrimination between adjacent Metavir F was very significant, especially between F1 and F0 or F2 (Table S15) as also previously suggested by AUROCs. In contrast, NITs were unable to discriminate Metavir F0 from F1 (Table S16) and their Obuchowski index was significantly lower than that of observer  $F_R$  in post-analysis group 2A (details in Supplemental Material).

### **Comparison with senior radiologists**

The accuracies of the single evaluations (on raw images of 10 patients) by the three senior evaluator radiologists were compared to those of the two evaluations of observers in group 2A (Table 4, Figure S4). Baseline HGE observer accuracies for fibrosis staging and cirrhosis diagnosis were not significantly different from radiologist accuracies except for fibrosis staging, which was significantly ( $p<0.001$ ) higher in expert 1 (full-time in HGE for clinical practice). However, post-analysis accuracies for HGE observers became higher to those of non-expert for fibrosis staging ( $p=0.014$ ) and cirrhosis ( $p=0.069$ ) and not significantly different from that of expert 1 for fibrosis staging ( $p=0.215$ ).

## DISCUSSION

**Main results** - Using a simple and original classification of liver fibrosis by CT, the accuracy for cirrhosis (84%) and discrimination for marked fibrosis (AUROC=0.83) by observers were already very satisfactory at baseline. Image analysis, especially on annotated images with simple descriptors, significantly improved accuracy for fibrosis staging or cirrhosis, as well as discrimination between fibrosis stages or for all binary diagnostic targets, and calibration of radiological classification. For fibrosis staging and cirrhosis diagnosis, the best progression was observed in group 2A (annotated abnormalities) (Figure 2A). Thus, simple annotations provided the greatest improvement for fibrosis staging by CT, which consequently became accurate. Thus, NPV and PPV for marked fibrosis ( $F \geq 3$ ) were 98% and 95%, respectively. Semi-automated morphometric analysis tended to improve the diagnostic performance of fibrosis staging but less so than did the annotated images in the study conditions. Finally, we observed high discrimination of all Metavir F by radiological  $F_R$  classification on annotated images, which must be clarified as it was unexpected.

**Originalities** - The Metavir F stages 0 and 1 were not distinguished by definition in the  $F_R$  classification. Metavir F was an independent predictor of  $F_R$  accuracy. Thus, accuracy was the highest in F4 and the lowest in F1 which was expected due to the preliminary merging of F0 and F1 in radiological staging. Unexpectedly, the overall (i.e. mean)  $F_R$  observer staging significantly discriminated F1 from F0 or F2 (Figure 3) in post-analysis group 2A despite the merging of  $F_R0$  and  $F_R1$  in the present study. This discrimination is noteworthy since NITs are usually unable to distinguish F0 from F1 [18] as in the present study, even with magnetic resonance elastography [19]. These data suggest that the radiological classification can be improved. Thus, some minor radiological abnormalities can distinguish Metavir F1 from F0, and early fibrosis may be detectable by CT. The high discrimination of radiological staging

finds reflection in the very high Obuchowski index at 0.873—significantly higher than that of NITs—in post-analysis group 2A. In comparison, with the best-performing tests in NAFLD, the Obuchowski index was, liver stiffness: 0.808, FibroMeter<sup>VCTE2G</sup>: 0.846 [20].

Accuracy for cirrhosis diagnosis was as high in Metavir F $\leq$ 3 (96%) as it was in F4 (98%) in post-analysis group 2A (Table S11), resulting in a very high NPV (99%) and a high PPV (85%) (Table S12). This is also noteworthy, since VCTE, still considered as the best-performing commonly-employed NIT for cirrhosis, provides a PPV of only around 60% [21]. Indeed, NITs more frequently classify cirrhosis correctly in patients with F $\leq$ 3 than they do in patients with F4. Because F4 is infrequent (10-15%), overall accuracy remains good despite a high rate of misclassified F4. Remarkably therefore, CT can avoid this usual pitfall. Finally, agreement between Metavir F, as reference, and F<sub>R</sub> stagings was excellent in post-analysis group 2A, with an ICC=0.95, in contrast to Metavir F agreement between expert, as reference, and first line pathologists as shown in Figure S5, with an ICC=0.87, or with an ICC=0.69 reported in a previous study [22].

**Interpretation** - In our prospective randomized stratified study, the absence of significant change in the post-analysis accuracy of the control group, with a classical presentation of semiology and the same images in the two steps, suggests good overall reproducibility and argues against a learning phenomenon attributable to the study. Thus, the control group validates the methodology used in our study. The excellent correlation between accuracy and observer agreement further validates the evaluation method. Unexpectedly, we observed only a weak improvement in accuracy with morphometry, which could seem paradoxical given the high level of information provided by it. Indeed, morphometry improved diagnostic accuracy in our preliminary study [6]. We hypothesise that the information presented for this technique may have been too complex for the observers. By contrast, the simple presentation including annotated images showed the best improvement in accuracy. The trivial explanation is that the

commented analysis makes it possible to visualize the anomalies easily and quickly and thus stage disease with greater accuracy. This good result was obtained despite the very novel character of the radiological classification in fibrosis stages, furthermore presented to naive observers. In contrast to fibrosis staging, there was a small but significant accuracy improvement for cirrhosis diagnosis with both image analyses. The moderateness of the improvement was attributable to the already very good baseline accuracy (with a limited progression margin), itself attributable to the classical character of this semiology and the presence of more marked anatomical abnormalities such as signs of portal hypertension. Considering the centre effect, there was mainly a significant effect in centre 5 of group 3B with a worsening in fibrosis staging (Figure 2D). This might be attributable to the combination of the complexity of morphometric data and poorer initial training. Indeed, this centre was the only one including three residents with unusual university training before internship. Finally, image analysis with annotations (group 2A) provided excellent inter-centre reproducibility for accuracy in both fibrosis staging and cirrhosis (Figure 2B/C). The present CT method is simpler and more discriminant than texture analysis [4]. However, a next step will be to compare the present modality of fibrosis staging with radiomics which are promising techniques [23, 24]. Radiomics have been evaluated in liver fibrosis staging with CT used alone [25] or combined with biomarkers [26], with magnetic resonance imaging [27] and with ultrasound combined with elastometry [28]. Therefore, radiomics should be compared with our CT technique which might include two advantages: easy applicability and discrimination between F0 and F1. The radiation doses should also be compared.

**Limitations** - The limits of this study are mainly methodological, concerning observer choice and image presentation with annotations. The present exploratory study might also include an optimism bias. These limits are detailed in the Supplemental material. One limitation of CT is radiation exposure, especially in patients with recurrent CT imaging [29]. However, this limit

can be circumvented or attenuated. First, this limit does not concern other initial validated indications of CT than liver fibrosis staging. For example, cirrhosis is often fortuitously discovered. Second, CT indications can be restricted to patients with discordant NITs for liver fibrosis, or when ultrasound imaging is not complete or inconclusive. Finally, one can use a low dose CT technique [30] or new devices reducing the radiation dose by 81% [31].

**Application** - Our study shows that annotated images improve diagnoses by HGE clinicians. We believe that this improvement could also concern other physicians as suggested by the accuracy difference between expert and non-expert radiologists (Table 4). Thus, the approach presented here can be applied in clinical practice via three means: 1) a diagnostic grid including the elementary descriptors, especially for first-line diagnoses (Figure 1A); 2) annotations provided by radiologists or by 3) semi-automatic morphometric analysis for second-line diagnoses. A reading grid contributes to structuring diagnostic reasoning. Furthermore, semi-automatic morphometry makes it possible to imagine, in the long term, an application to clinical practice not only as a diagnostic aid for the radiologist evaluating images initially, but also as a means for clinician prescribers and patients themselves to better appropriate results. Upstream of these possibilities, it will be necessary to validate this new semiology (especially for fibrosis  $F_R$  staging into five stages) with comparisons to a robust reference (liver biopsy, NITs) in a large patient population. Further study will determine independent descriptors and an ensuing diagnostic score. Finally, in the future, a single radiological modality could include a comprehensive liver evaluation comprising fibrosis staging, elastometry, steatosis and masses.

**Conclusion** - A new CT classification based on simple descriptors can accurately stage liver fibrosis according to the Metavir system. Unexpectedly, this classification, although based on cirrhosis descriptors, was able to discriminate between all Metavir fibrosis stages, including between F0 and F1. The simple identification of these descriptors was superior to

morphometric analysis but clinical application needs to be improved. This simple descriptor identification might result in accuracies by HGE clinicians at least as good as those provided by senior radiologists. Finally, imaging could provide comprehensive liver evaluation including highly discriminative liver fibrosis staging.

**Acknowledgments:**

Bouguen Guillaume, inter-regional coordinator of HGE resident teaching in the West France  
Lerolle Nicolas, Dean of Medicine Faculty of Angers; Annweiler Cédric, Director of the  
Medicine Department of Medicine Faculty of Angers; Chrétien Jean Marie for regulatory  
aspects; Erwin Kevin L. for writing assistance (English proofreading); and the residents in  
HGE:

**CDMIR group:** Gieu Marie, Griveaux Océane, Imbert Astrid, Le Bras Pierrine, Rochereau  
Antoine, Stephant Sophie, Gouriou Claire, Laland Mathilde, Michaux Quentin, Moisset  
Laure, Quenin Benoît, Sarfati Laura, Thomas Quentin, Planade Oriane, Reboux Noémi,  
Tanneche Jérémy, Texier Charles, Verdier Marine, Wilson Audrey, Marone Dior, Marraud  
des Grottes Hortense, Noel Catherine, Broussier Thomas, Fozou Djeutchouang Sfoz Corine,  
Candillier Camille, Canevet Mickaël, David Nicolas, Denjean Théo, Dincuff Eloïse, Fleurent  
Coralie, Chautard Romain, Cornu Isabelle, Duval Lucie, Fontana Julien, Garcia Géraldine-  
Anne, Gondran Hannah, Esnaud Élise, Grainville Thomas, Julien Hélène, Lot Alix,  
Lucidarme Camille, Merle Pauline, Moussaddaq Anne-Sophie, Jaillais Anais, Lahlou Widad,  
Landrieux Amandine, Lepicard Julia, Baleba Thomas, Balzano Vittoria, Boileve Juliette,  
Bourgeois Geoffrey, Brunet Tanguy, Alessa Fahed, Barbin Eva, Boisteau Emeric, Bordet  
Martin, Bresson Hélène, Cadieu Estelle, Azarfane Miloud, Buisson François, Charles Anna,  
Cherid Karima, Daubigny Augustin, De Maissin Astrid, Desgrange Clarisse, Douillard  
Marion, El Fettouhi Abdessamad, Elizondo Mélodie, Herissay Auguste, Nana Sede, Mbakop  
Raissa and Mekkassi Mohammed.

## REFERENCES

- [1] Bedossa P, Carrat F. Liver biopsy: the best, not the gold standard. *J Hepatol* 2009;50(1):1-3.
- [2] European Association for the Study of the L, List of panel m, Berzigotti A, Boursier J, Castera L, Cazzagon N, et al. EASL Clinical Practice Guidelines On Non-Invasive Tests For Evaluation Of Liver Disease Severity And Prognosis- 2021 Update. *J Hepatol* 2021.
- [3] Laghi A, Iannaccone R, Catalano C, Carbone I, Ferrari R, Mangiapane F, et al. Multiphase multislice spiral CT for liver assessment: optimization in cirrhotic patients. *Radiol Med* 2002;103(3):188-95.
- [4] Budai BK, Toth A, Borsos P, Frank VG, Shariati S, Fejer B, et al. Three-dimensional CT texture analysis of anatomic liver segments can differentiate between low-grade and high-grade fibrosis. *BMC Med Imaging* 2020;20(1):108.
- [5] Rutjes AW, Reitsma JB, Coomarasamy A, Khan KS, Bossuyt PM. Evaluation of diagnostic tests when there is no gold standard. A review of methods. *Health Technol Assess* 2007;11(50):iii, ix-51.
- [6] Cales P, Lebigot J, Chaigneau J, Boursier J, Paisant A, Oberti F, et al. Morphometry on CT improves non-invasive diagnosis of hepatic fibrosis. <https://afefasso.fr/wp-content/uploads/2018/10/CO-AFEF-2018pdf> 2018.
- [7] Aube C. Imaging modalities for the diagnosis of hepatic fibrosis and cirrhosis. *Clin Res Hepatol Gastroenterol* 2015;39(1):38-44.
- [8] Indiran V, Jagannathan K. Right posterior hepatic notch sign. *Abdom Radiol (NY)* 2017;42(11):2780.



- [9] Aube C, Oberti F, Korali N, Namour MA, Loisel D, Tanguy JY, et al. Ultrasonographic diagnosis of hepatic fibrosis or cirrhosis. *J Hepatol* 1999;30(3):472-8.
- [10] Kudo M, Zheng RQ, Kim SR, Okabe Y, Osaki Y, Iijima H, et al. Diagnostic accuracy of imaging for liver cirrhosis compared to histologically proven liver cirrhosis. A multicenter collaborative study. *Intervirol* 2008;51 Suppl 1:17-26.
- [11] Intraobserver and interobserver variations in liver biopsy interpretation in patients with chronic hepatitis C. The French METAVIR Cooperative Study Group. *Hepatology* 1994;20(1 Pt 1):15-20.
- [12] Cales P, Chaigneau J, Hunault G, Michalak S, Cavaro-Menard C, Fasquel JB, et al. Automated morphometry provides accurate and reproducible virtual staging of liver fibrosis in chronic hepatitis C. *J Pathol Inform* 2015;6:20.
- [13] Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig L, et al. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. *BMJ* 2015;351:h5527.
- [14] Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *BMJ* 2015;350:g7594.
- [15] Boursier J, de Ledinghen V, Poynard T, Guéchet J, Carrat F, Leroy V, et al. An extension of STARD statements for reporting diagnostic accuracy studies on liver fibrosis tests: The Liver-FibroSTARD standards. *J Hepatol* 2015;62(4):807-15.
- [16] Lambert J, Halfon P, Penaranda G, Bedossa P, Cacoub P, Carrat F. How to measure the diagnostic accuracy of noninvasive liver fibrosis indices: the area under the ROC curve revisited. *Clin Chem* 2008;54(8):1372-8.

- [17] Katz MH. Multivariate analysis. A practical guide for clinicians. Cambridge university press 1999.
- [18] European Association for the Study of the Liver, Asociacion Latinoamericana para el Estudio del Hgado. EASL-ALEH Clinical Practice Guidelines: Non-invasive tests for evaluation of liver disease severity and prognosis. *J Hepatol* 2015;63(1):237-64.
- [19] Hsu C, Caussy C, Imajo K, Chen J, Singh S, Kaulback K, et al. Magnetic Resonance vs Transient Elastography Analysis of Patients With Nonalcoholic Fatty Liver Disease: A Systematic Review and Pooled Analysis of Individual Participants. *Clin Gastroenterol Hepatol* 2019;17(4):630-7 e8.
- [20] Ducancelle A, Leroy V, Vergniol J, Sturm N, Le Bail B, Zarski J, et al. A single test combining blood markers and elastography is more accurate than other fibrosis tests in the main causes of chronic liver diseases *J Clin Gastro* 2017;51(7):639-49.
- [21] Cales P, Boursier J, Oberti F, Bardou D, Zarski JP, de Ledinghen V. Cirrhosis diagnosis and liver fibrosis staging: transient elastometry versus cirrhosis blood test. *J Clin Gastroenterol* 2015;49(6):512-9.
- [22] Rousselet MC, Michalak S, Dupre F, Croue A, Bedossa P, Saint-Andre JP, et al. Sources of variability in histological scoring of chronic viral hepatitis. *Hepatology* 2005;41(2):257-64.
- [23] Wei J, Jiang H, Gu D, Niu M, Fu F, Han Y, et al. Radiomics in liver diseases: Current progress and future opportunities. *Liver Int* 2020;40(9):2050-63.
- [24] Decharatanachart P, Chaiteerakij R, Tiyyarattanachai T, Treeprasertsuk S. Application of artificial intelligence in chronic liver diseases: a systematic review and meta-analysis. *BMC Gastroenterol* 2021;21(1):10.

- [25] Cui E, Long W, Wu J, Li Q, Ma C, Lei Y, et al. Predicting the stages of liver fibrosis with multiphase CT radiomics based on volumetric features. *Abdom Radiol (NY)* 2021;46(8):3866-76.
- [26] Hu P, Hu X, Lin Y, Yu X, Tao X, Sun J, et al. A Combination Model of Radiomics Features and Clinical Biomarkers as a Nomogram to Differentiate Nonadvanced From Advanced Liver Fibrosis: A Retrospective Study. *Acad Radiol* 2021.
- [27] Zheng R, Shi C, Wang C, Shi N, Qiu T, Chen W, et al. Imaging-Based Staging of Hepatic Fibrosis in Patients with Hepatitis B: A Dynamic Radiomics Model Based on Gd-EOB-DTPA-Enhanced MRI. *Biomolecules* 2021;11(2).
- [28] Xue LY, Jiang ZY, Fu TT, Wang QM, Zhu YL, Dai M, et al. Transfer learning radiomics based on multimodal ultrasound imaging for staging liver fibrosis. *Eur Radiol* 2020;30(5):2973-83.
- [29] Kwee TC, Dijkstra H, Knapen DG, de Vries EGE, Yakar D. Which patients are prone to undergo disproportionate recurrent CT imaging and should we worry? *Eur J Radiol* 2020;125:108898.
- [30] Takahashi H, Okada M, Hyodo T, Hidaka S, Kagawa Y, Matsuki M, et al. Can low-dose CT with iterative reconstruction reduce both the radiation dose and the amount of iodine contrast medium in a dynamic CT study of the liver? *Eur J Radiol* 2014;83(4):684-91.
- [31] Leyendecker P, Faucher V, Labani A, Noblet V, Lefebvre F, Magotteaux P, et al. Prospective evaluation of ultra-low-dose contrast-enhanced 100-kV abdominal computed tomography with tin filter: effect on radiation dose reduction and image quality with a third-generation dual-source CT system. *Eur Radiol* 2019;29(4):2107-16.

## Figure legend

**Figure 1. Liver CT images.** Panel A: Reading grid for liver fibrosis staging. Panels B to F: Examples of evaluated images according to evaluation steps and observer groups. Panel B shows automatic measurements of liver (red) and spleen (magenta) surfaces, allowing volume calculation by successive CT slices as reported in patient #10 (Metavir F2) of [Table S2](#). The yellow pixels indicate fat contained in the box drawn around the liver to calculate the atrophy index (ratio between the fat area in the box and the box area). Panel C shows the coarse aspect and main morphometric measurements as reported in patient #7 (Metavir F3) of [Table S2](#).

**Figure 2. Observer  $F_R$  accuracies.** Panel A: Accuracies as a function of observer groups with p values between evaluation steps. Panels B/C: Accuracies for fibrosis staging (B) and cirrhosis (C) as a function of centres, observer groups, and evaluation steps. Panel D: Accuracy course (net change) as a function of observer groups and centres. \* indicates an interaction between centre and observer group; # indicates a significant centre effect.

**Figure 3. Agreement of radiological  $F_R$  classification vs Metavir F.** Agreement as a function of evaluation steps and observer groups. At baseline, calibration was poor in F0 and F1. In post-analysis evaluation, the calibration became excellent for all F in group 2A.  $r_s$ : Spearman's correlation coefficient for individual observer data; comparisons are detailed in [Table S14](#).

**Table 1.** Study design: the four steps of the evaluation session.

	Observer group		
	1C	2A	3B
Images tested	Raw (Fig. 1E)	Annotated (Fig. 1D/F)	Morphometric analysis (Fig. 1B/C)
Step:			
1. <i>Study presentation</i> ( <i>briefing</i> )	Introduction 18 images with comments on the 7 descriptors <sup>a</sup>		
2. <i>Baseline evaluation</i> <sup>b</sup>	Raw images		
3. <i>Explanation on evaluated</i> <i>semiology</i>	Classical semiology		Morphometry
4. <i>Post-analysis evaluation</i> <sup>bc</sup>	Raw images	Annotated images <sup>d</sup>	Images with morphometry <sup>e</sup>

<sup>a</sup> Example in [Figure 1D](#). Patients different from those of the next evaluation steps.

<sup>b</sup> Patient order identical in the 3 groups.

<sup>c</sup> Patient order different from step #2.

<sup>d</sup> 35 comments including at least each of the seven categorized descriptors of liver fibrosis/cirrhosis.

<sup>e</sup> With data as in [Table S2](#).

**Table 2.** Comparison of F<sub>R</sub> accuracies (%) between groups.

	Observer group				Comparison ( <i>p</i> )			
	All	1C	2A	3C	All <sup>a</sup>	1 vs 2 <sup>b</sup>	1 vs 3 <sup>b</sup>	2 vs 3 <sup>b</sup>
Measures per step (n)	700	220	210	270	700	430	490	480
<b>A. Fibrosis F<sub>R</sub> stages:</b>								
Accuracy (%):								
Baseline	43.1	45.0	46.7	38.9	0.186	0.772	0.197	0.094
Post-analysis	60.4	50.9	81.0	52.2	<0.001	<0.001	0.786	<0.001
<i>p</i> <sup>c</sup>	<0.001	0.187	<0.001	0.001	-	-	-	-
<i>Kappa</i>	0.225	0.247	0.250	0.164	-	-	-	-
Course (%) <sup>d</sup> :								
Stable	60.1	62.3	61.0	57.8	<0.001	<0.001	0.317	<0.001
Regression	11.3	15.9	2.4	14.4	0.576	0.843	0.354	0.513
Progression	28.6	21.8	36.7	27.8	<0.001	<0.001	0.704	<0.001
Net change (%) <sup>e</sup>	17.3±60.8	5.9±61.3	34.3±52.4	13.3±63.7	0.001	0.001	0.143	0.048
<b>B. Cirrhosis:</b>								
Accuracy (%):								
Baseline	84.1	81.8	88.1	83.0	0.163	0.080	0.811	0.122
Post-analysis	89.3	81.8	96.3	90.0	<0.001	<0.001	0.012	0.012
<i>p</i> <sup>c</sup>	0.001	1	<0.001	0.005	-	-	-	-
<i>Kappa</i>	0.334	0.419	0.132	0.326	-	-	-	-
Course (%) <sup>d</sup> :								
Stable	84.6	82.7	87.1	84.1	0.044	0.017	0.113	0.436
Regression	5.1	8.6	2.4	4.4	0.430	0.226	0.715	0.364
Progression	10.3	8.6	10.5	11.5	0.011	0.006	0.064	0.320
Net change (%) <sup>e</sup>	5.1±39.0	0.0±41.7	8.1±35.0	7.0±39.4	0.584	0.623	0.368	0.771
<b>C. Comparison of fibrosis F<sub>R</sub> stages vs cirrhosis (<i>p</i><sup>e</sup>):</b>								
Baseline accuracy	<0.001	<0.001	<0.001	<0.001	-	-	-	-
Post-analysis accuracy	<0.001	<0.001	<0.001	<0.001	-	-	-	-
Course:								
Stable	<0.001	<0.001	<0.001	<0.001	-	-	-	-
Regression	<0.001	0.011	1	<0.001	-	-	-	-
Progression	<0.001	<0.001	<0.001	<0.001	-	-	-	-
Net change	<0.001	0.134	<0.001	0.091	-	-	-	-

<sup>a</sup> Chi<sup>2</sup> or ANOVA.

<sup>b</sup> Unpaired Fisher test or Student *t* test

<sup>c</sup> Paired McNemar test for difference between baseline and post-analysis evaluation (i.e. net change) or McNemar-Bowker test.

<sup>d</sup> Cumulative rate of observer change per patient (expressed in binary qualitative value). This rate is therefore different from the changes in mean accuracy for 10 patients depicted in the main text (Table S4).

<sup>e</sup> The net change is the absolute difference between baseline and post-test analysis and is expressed as a quantitative variable.

**Table 3.** Observer  $F_R$  AUROC for binary diagnostic targets of Metavir F as a function of observer groups and evaluation steps.

Target	Group						All
	1C	2A	$p$ vs 1C <sup>a</sup>	3B	$p$ vs 1C <sup>a</sup>	$p$ vs 2A <sup>a</sup>	
Measures (n)	220	210	-	270	-	-	700
<b>F<math>\geq</math>1:</b>							
Baseline	0.811	0.758	0.258	0.717	<b>0.048</b>	0.427	0.759
Post-analysis	0.760	0.894	<b>0.003</b>	0.797	0.452	<b>0.010</b>	- <sup>b</sup>
$p$ <sup>c</sup>	0.185	<b>&lt;0.001</b>	-	0.072	-	-	-
<b>F<math>\geq</math>2:</b>							
Baseline	0.689	0.665	0.647	0.621	0.170	0.505	0.655
Post-analysis	0.698	0.933	<b>&lt;0.001</b>	0.754	0.232	<b>&lt;0.001</b>	- <sup>b</sup>
$p$ <sup>c</sup>	0.780	<b>&lt;0.001</b>	-	<b>&lt;0.001</b>	-	-	-
<b>F<math>\geq</math>3:</b>							
Baseline	0.855	0.858	0.935	0.801	0.150	0.128	0.834
Post-analysis	0.856	0.987	<b>&lt;0.001</b>	0.843	0.708	<b>&lt;0.001</b>	- <sup>b</sup>
$p$ <sup>c</sup>	0.967	<b>&lt;0.001</b>	-	0.155	-	-	-
<b>F=4:</b>							
Baseline	0.834	0.901	0.107	0.836	0.964	0.084	0.854
Post-analysis	0.861	0.976	<b>&lt;0.001</b>	0.927	<b>0.034</b>	<b>0.013</b>	- <sup>b</sup>
$p$ <sup>c</sup>	0.409	<b>0.008</b>	-	<b>0.001</b>	-	-	-
<b>All F<sup>d</sup></b>							
Baseline	0.672	0.658	<b>&lt;0.001</b>	0.621	<b>&lt;0.001</b>	<b>&lt;0.001</b>	0.648
Post-analysis	0.677	0.873	<b>&lt;0.001</b>	0.715	<b>&lt;0.001</b>	<b>&lt;0.001</b>	- <sup>b</sup>
$p$	0.854	<b>&lt;0.001</b>	-	<b>&lt;0.001</b>	-	-	-

Significant differences in bold.

<sup>a</sup> Unpaired reduced gap test.

<sup>b</sup> Not calculated because not appropriate due to group 1C.

<sup>c</sup> Paired Delong test.

<sup>d</sup> Discrimination by Obuchowski index and test.

**Table 4.** Comparison of accuracies (%) between group 2A observers and radiologists as a function of evaluation steps.

	Observers		Radiologists					
	Baseline	Post-analysis	$p^a$ vs baseline observer		$p^b$ vs post-analysis observer		$p^c$ vs 10 patients by radiologist	
Patients (n)	10	10	10	-	-	96	-	106
						<b>Expert 1</b>		
F <sub>R</sub> staging	47	81	100	<b>&lt;0.001</b>	0.215	51	<b>0.002</b>	56
Cirrhosis	88	96	100	0.608	1	86	0.607	88
						<b>Expert 2</b>		
F <sub>R</sub> staging	47	81	60	0.523	0.116	41	0.318	42
Cirrhosis	88	96	100	0.609	1	83	0.353	85
						<b>Non-expert</b>		
F <sub>R</sub> staging	47	81	20	0.116	<b>0.014</b>	32	0.721	31
Cirrhosis	88	96	80	0.353	0.069	78	1	78

Significant differences in bold.




















<sup>a</sup> Comparison of radiologist vs observer baseline accuracy in the 10 patients of the present study by Fisher test.

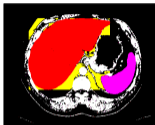
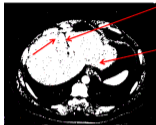
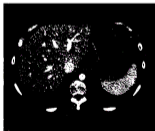
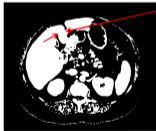
<sup>b</sup> Comparison of radiologist vs observer post-analysis accuracy in the 10 patients of the present study by Fisher test.

<sup>c</sup> The radiologist comparison was performed between accuracies measured in the 10 patients of the core population vs the remaining 96 patients evaluated by radiologists <sup>8</sup>.



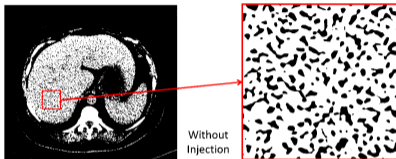
### Reading grid for liver fibrosis staging by computed tomography

Aspect	Normal	Moderate	Marked	Notation
Liver edge irregularity				
Segment I hypertrophy				
Segment IV hypotrophy				
Coarse appearance of liver parenchyma				
Portosystemic collaterals				
<b>Measurement</b>				
Spleen size (major axis)				
<b>Abnormality</b>	<b>Absent</b>		<b>Present</b>	
Right posterior notch				

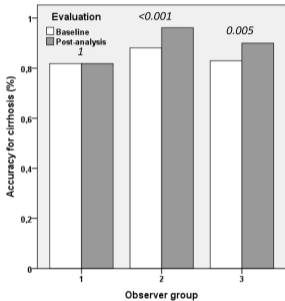
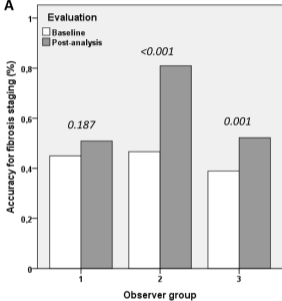
**B****Step 4 group 3**Images with  
morphometric  
analysis**D****Step 1 all groups**Marked  
hypotrophy of  
segment IVHypertrophy of  
left lobe**E****Step 2 all groups. Step 4 group 1**Non-annotated  
images**F****Step 4 group 2**Marked  
hypotrophy of  
segment IV

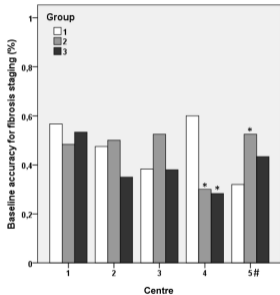
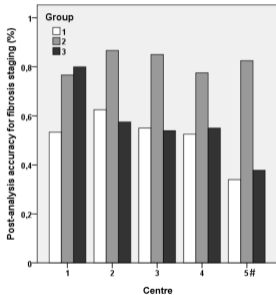
Commented images

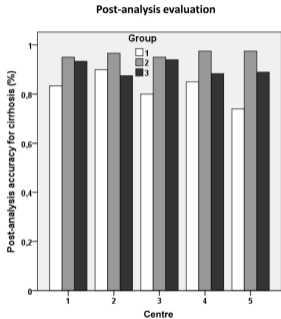
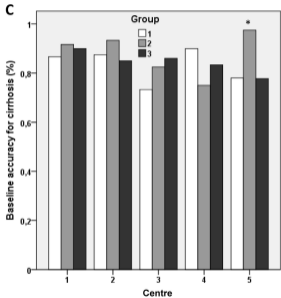
C Images with morphometric analysis: step 4 in group 3 & step 5 in groups 1 & 2

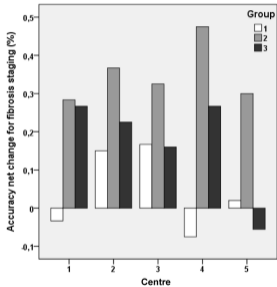
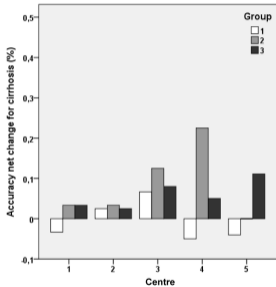


CT score for Metavir F	CT score for cirrhosis	Segment I surface (cm <sup>2</sup> )	Total liver density (median HU)	Spleen volume (cm <sup>3</sup> )	Maximum liver perimeter (cm)	Atrophy index	Furrow thickness (cm)
3.71	0.1996	9.01	137	271	79.9	11.69	0.6

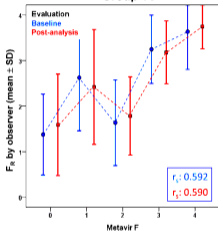
**A**

**B****Baseline evaluation****Post-analysis evaluation**

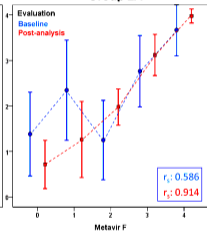


**D****Baseline evaluation****Post-analysis evaluation**

Group 1C



Group 2A



Group 3B

