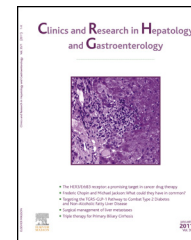Available online at

ScienceDirect
www.sciencedirect.com

Elsevier Masson France

EM|consulte
www.em-consulte.com/en

Clinics and Research in Hepatology and Gastroenterology

ELSEVIER MASSON

suppl. Informations

ORIGINAL ARTICLE

Learning curve and intra/interobserver agreement of transient elastography in chronic hepatitis C patients with or without HIV co-infection

Hugo Perazzo [a,*], Flavia F. Fernandes [b], João Carlos Soares [a], Juliana Fittipaldi [a], Sandra W. Cardoso [a], Beatriz Grinsztejn [a], Valdiléa G. Veloso [a]

[a] National Institute of Infectious Disease - Fundação Oswaldo Cruz, Laboratory of clinical research on STD/AIDS, Rio de Janeiro, Brazil
[b] Bonsucesso Federal Hospital, Gastroenterology Department, Rio de Janeiro, Brazil

Summary

Background and objectives: Liver stiffness measurement (LSM) by transient elastography has been validated as a noninvasive method to stage liver fibrosis. Few studies have evaluated the learning curve of this method and its reproducibility has led to controversy results. We aimed to evaluate the intra- and interobserver agreement of transient elastography as well as its learning curve for definition of an experimented operator.

Methods: We retrospectively analyzed 922 examinations performed in 544 patients during a training program of transient elastography. Patients with chronic hepatitis C with or without HIV co-infection that had two examinations by the training operator (intraobserver analysis; $n = 125$) or examination by both training and experimented operators (interobserver analysis; $n = 151$) in the same day were included. LSM was converted to METAVIR score: < 7.1 as F0F1, $7.1—9.4$ as F2, $9.5—12.4$, as F3 and > 12.4 kPa as F4.

Results: The overall intra- and interobserver intraclass correlation coefficient [ICC 95% CI] were 0.926 $(0.901—0.951)$ and 0.912 $(0.885—0.939)$, respectively. Measurements were correlated [Spearman's] in intra- $[0.906, P < 0.0001]$ and interobserver $[0.907, P < 0.0001]$ analysis. Reliability values [kappa (SE)] were $k = 0.74$ (0.09) and $k = 0.85$ (0.08) for fibrosis stages $F \geq 2$ and $k = 0.77$ (0.09) and $k = 0.75$ (0.08) for cirrhosis in intra- and interobserver analysis, respectively.

Agreement was improved when operator's experience was higher than 100 exams. However, it was observed discordance for fibrosis staging between examinations in a quarter of patients.
Conclusion: Although there was a considerable discrepancy on fibrosis staging between examinations and a small power, transient elastography had an acceptable reproducibility in our population. Performance of at least 100 examinations should be used to define an experimented operator.

Introduction

Chronic hepatitis C remains one of the main causes of chronic liver disease worldwide. Co-infection by human immunodeficiency virus (HIV) and hepatitis C virus (HCV) are highly prevalent [1,2]. HIV/HCV infection has been associated with more aggressive liver disease and higher mortality compared to HCV mono-infection [2]. The correct staging of liver fibrosis has important implications for prognostic, therapeutic and monitoring purposes in chronic hepatitis C [3]. Transient elastography by FibroScan® (EchoSens, Paris, France), based on liver stiffness measurement (LSM), has been validated as an accurate noninvasive method to stage liver fibrosis in chronic liver diseases [4]. However, this method has been challenged by its feasibility and controversies results on intra- and interobserver variability [5—7]. Currently, the performance of at least 500 examinations has been empirically established to define an experimented operator [8]. However, to date no study has properly evaluated the learning curve of this method.

The aims of our study were:

- to evaluate the intra- and interobserver agreement of transient elastography obtained in the same day in patients with chronic hepatitis C with or without HIV co-infection;
- to evaluate the learning curve of an initially inexperience operator in transient elastography and determine the minimal number of examinations needed to define an experimented operator.

Material and methods

Study design

This cross-sectional study was conducted on the National Institute of Infectious Diseases (INI) at Fundação Oswaldo Cruz (FIOCRUZ), Rio de Janeiro, Brazil. The database of all transient elastographies performed from February 2012 to April 2014 was retrospectively analyzed to assess the learning curve and variability of this method. During this period an initially inexperienced operator on transient elastography was training at the referring center. Consecutive patients older than 18 years, having chronic infection by hepatitis C with or without HIV co-infection and submitted to transient elastography twice by the training operator or by both training and experimented operators at the same day were included. The exclusion criteria were other liver disease than chronic hepatitis C and unreliable liver stiffness measurement in at least one examination. Laboratory blood analysis was realized in a maximal delay of 6 months before or after transient elastography. The study protocol was conducted in accordance with the 1975 Helsinki declaration as revised in 1983, and all patients have given a consent allowing a retrospective analysis of their exams results.

Program training on transient elastography

During the period of recruitment for this analysis a training operator (JCS) on transient elastography was under supervision of an experimented operator (FFF) until achievement of an experience of more than 500 exams [8]. The training operator had an extensive formation on theory before starting the performance of the transient elastography examinations. The training program consisted on:

- visual supervision by the experimented operator;
- repetition of the examination twice in the same patient at the same day by the training operator in few randomly chosen patients;
- repetition of the examination by the experimented operator afterwards the training operator in the same patient at the same day for confirmation of results in few randomly chosen patients.

The learning curve of transient elastography was based on intra- and interobserver agreement taking into account the training operator experience measured in number of examinations. The experience of the training operator was considered as poor (< 100 exams), moderate (101—300 exams), high (301—500 exams) and very high (> 500 exams). In case of discordance between results performed by the single training operator, both operators discussed to achieve a consensus for the final result. In patients with examinations by both operators, the result from the experimented operator was taken into account for patient's management.

Transient elastography

Transient elastography was performed following a validated procedure using an M probe of FibroScan® (EchoSens, Paris, France) [9]. Briefly, the patient was placed in dorsal decubitus with right arm in maximal abduction and the tip of the probe placed perpendicularly in an intercostal space at the level of the right lobe of the liver, usually where liver biopsy would be performed (i.e., intersection between the xiphoid
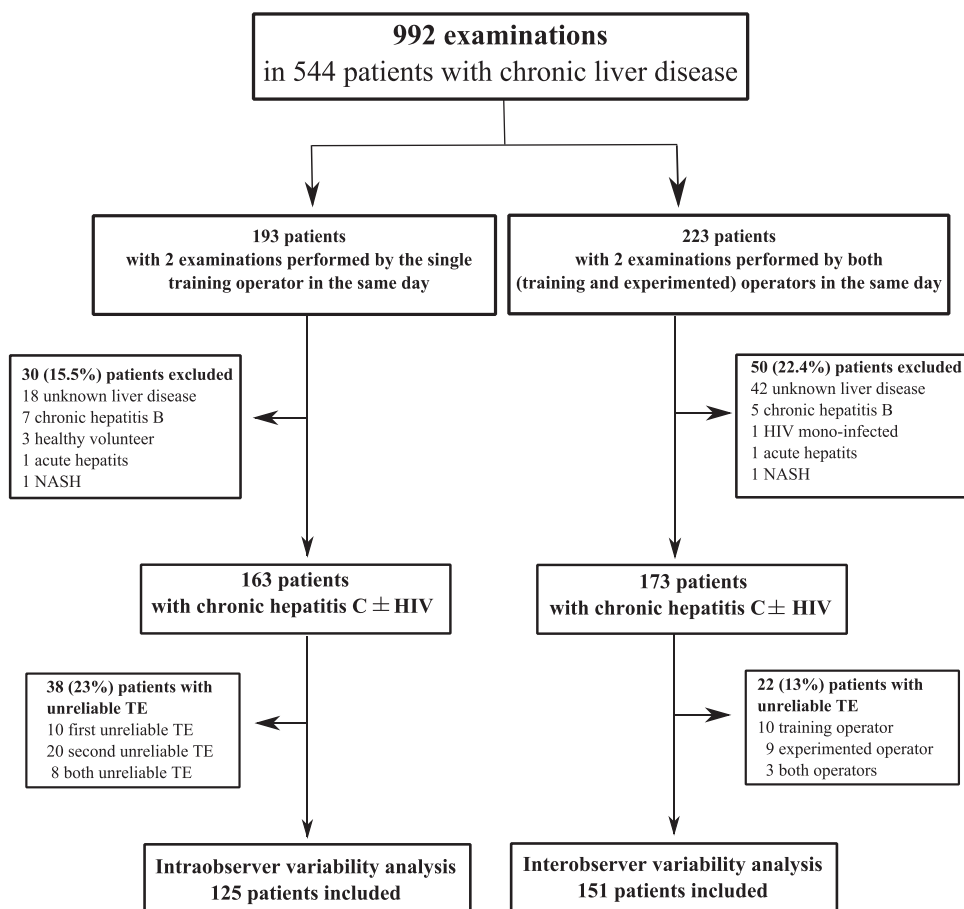
Figure 1 Study flow chart of patient's recruitment.

line and the median axillary line). The operator, assisted by a time-motion ultrasound image, located a liver portion free of large vascular structures where measures would not be influenced by respiratory phases and acquired a measurement. Transient elastography was considered reliable when the following criteria had been met:

- 10 successful measurements;
- an interquartile range lower than 30% of the median value;
- a success rate of more than 60% [10].

Liver stiffness was considered as the median of all valid measurements.

Liver stiffness measurement was defined as unreliable if did not meet the above-described criteria. Liver fibrosis, estimated by transient elastography, was converted to the METAVIR scoring system [11] as proposed by Castera et al. [12]: < 7.1 as F0F1; 7.1−9.4 as F2; 9.5−12.4 as F3 and > 12.4 kPa as F4.

In patients with repeated measurements, examinations were performed at the same day (one test just after the other), either by the training operator or by both training and experimented operators. Examinations were systematically performed by the training operator followed by the experimented operator in patients with the same fasting status. In the second examination, neither the training nor

the experimented operators were blinded to liver stiffness measurement performed in the first examination.

Statistical analysis

Continuous variables were reported as median (interquartile range). Discrete variables were reported as absolute and relative frequency. Comparisons between groups were assessed by non-parametric tests:

- Mann-Whitney whether two groups or Kruskal-Wallis whether more than two groups for quantitative comparisons;
- Fisher's exact test for qualitative comparisons.

Repetitive measures were compared and assessed by Wilcoxon signed-rank for paired continuous variables, McNemar for paired discrete variables and Spearman's rank correlation. Intraobserver agreement was analyzed in patients that had two examinations by the training operator and interobserver agreement in those that had examination by both operators (training and experimented) at the same day. Intra- and interobserver agreement was assessed using the intraclass correlation coefficient (ICC) and Kappa (κ) index. Significance level was determined when $P \leq 0.05$ assuming two-tailed tests. Statistical analyses

were performed using STATA statistical package for Windows (2012; StataCorp LP, College Station, TX, USA).

Results

During the recruitment period, training and experiment operators performed 992 examinations in 544 patients. Among those, 193 patients had two consecutive examinations performed by the training operator and 223 patients had repeated examination by the training and the experimented operator at the same day (Fig. 1).

Intraobserver analysis

Among the 193 eligible patients (a total of 386 examinations), 30 (16%) patients had no chronic hepatitis C and 38 (23%) had an unreliable transient elastography on at least one examination (Fig. 1). Thus, 125 patients [41% male gender, median (IQR) age 54 (46—62) years, 26% with HIV infection] were included for the intraobserver analysis (Table 1).

The overall intraobserver agreement ICC (95% confidence interval) was 0.926 (0.901—0.951) and the overall correlation between first and second examinations performed by the training operator was excellent (Spearman's value $\rho = 0.906$, $P < 0.0001$) (Table 2 and Fig. 2A). However, the intraobserver agreement was poor when the training operator had lower than 100 exams experience [ICC was 0.748 (0.504—0.993) and Spearman's value was 0.586, $P = 0.0353$] (Table 2). There was a clear improve on intraobserver agreement of the training operator related to experience increase in number of examinations over the training program. The agreement rates were higher when the training operator experience was at least moderate (> 100 exams) compared to poor (≤ 100 exams) [ICC: 0.930 (0.905—0.955) from 0.748 (0.504—0.993) and Spearman's value: 0.921 ($P < 0.0001$) from 0.586 ($P = 0.0353$)]. However, there was no difference on intraobserver agreement comparing the training operator experience as high (> 300) or not high (≤ 300) (Tables 2 and 3).

The overall intraobserver reliability k values (standard error) were, $\kappa = 0.74$ (0.09), $\kappa = 0.84$ (0.09) and $\kappa = 0.77$ (0.09) for fibrosis stages $F \geq 2$, for fibrosis stages $F \geq 3$ and for cirrhosis, respectively. The k values followed the same trend of intraobserver agreement according to the operator experience (Tables 2 and 3). Besides an excellent overall agreement, discordance of at least one stage and for two or more stages of fibrosis between first and second examinations was 24% and 6%, respectively.

Interobserver analysis

Among the 223 eligible patients (a total of 446 examinations), 50 (22%) patients were excluded due other liver disease than chronic hepatitis C and 22 (13%) due to unreliability of transient elastography of at least one operator (Fig. 1). Therefore, 151 patients [55% male gender, median (IQR) age 55 (45—61) years, 38% with HIV infection] were included for the interobserver analysis (Table 1).
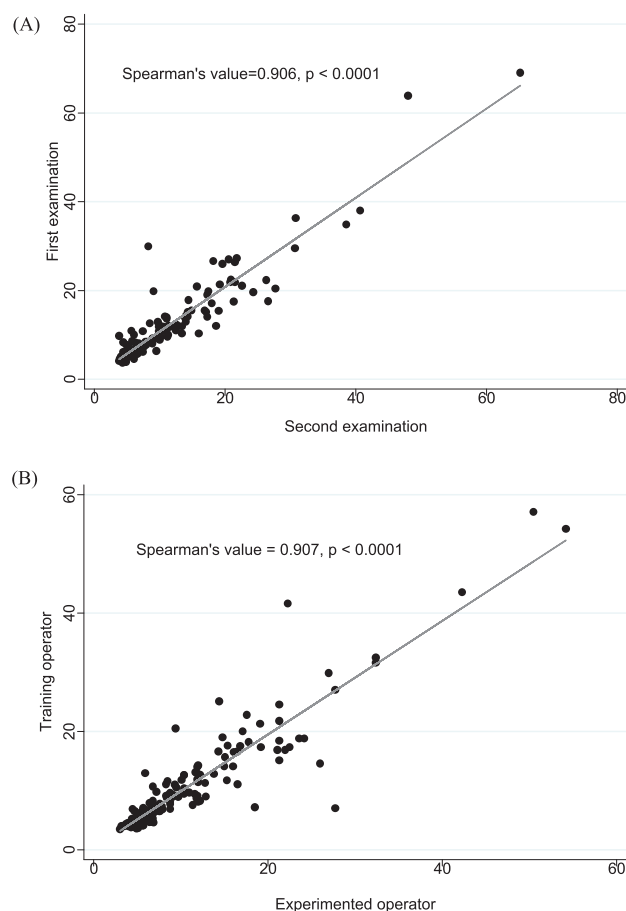

Figure 2 Graphic representation of the correlation between examinations in (A) intraobserver analysis and (B) interobserver analysis.

The overall intraobserver agreement ICC (95% confidence interval) was 0.912 (0.885—0.939) and the overall correlation between the training and experimented operators was also excellent (Spearman's value $\rho = 0.907$, $P < 0.0001$) (Table 2 and Fig. 2B). The overall interobserver reliability k values (standard error) were $\kappa = 0.85$ (0.08), $\kappa = 0.74$ (0.08) and $\kappa = 0.75$ (0.08) for fibrosis stages $F \geq 2$, for fibrosis stages $F \geq 3$ and for cirrhosis, respectively. Besides the good agreement rates, discordance of at least one stage and for two or more stages of fibrosis between examinations from training and experimented operators was observed in 25% and 4% of patients, respectively (Table 2). There was no increase on agreement rates related to training operator experience. Interobserver agreement rates were similar when training operator experience was poor (≤ 100 exams) or at least moderate (> 100 exams) (Tables 2 and 3).

Discussion

To date, only few studies have reported the variability of transient elastography and the learning curve of this method has not been evaluated. This study highlighted that 100 exams experience could be used to define an experimented operator with a satisfactory intra- and interobserver agreement of liver stiffness measurements by transient

Table 1 Characteristics of patients included in intra- and interobserver variability analysis.

	Intraobserver analysis (n = 125)	Interobserver analysis (n = 151)
Male gender	51 (41)	83 (55)
Age, years	54 (46–62)	55 (45–61)
HIV co-infection	32 (26)	58 (38)
HCV genotype-1	45 (87)	82 (75)
HCV treatment naive	33 (79)	56 (59)
Liver fibrosis, estimated by transient elastography		
F0F1, < 7.1 kPa	35 (28)	65 (43)
F2, 7.1–9.4 kPa	23 (18)	26 (17)
F3, 9.5–12.4 kPa	24 (19)	17 (11)
F4, > 12.4 kPa	43 (35)	43 (29)
Biological analysis		
ALT, U/L	55 (41–87)	59 (38–97)
AST, U/L	45 (33–68)	42 (30–71)
Alkaline Phosphatases, U/L	81 (67–111)	83 (66–112)
GGT, U/L	74 (42–141)	68 (41–110)
Total bilirubin, mg/dL	0.60 (0.45–1.00)	0.60 (0.37–0.80)
Triglycerides, mg/dL	107 (77–142)	104 (81–126)
Total cholesterol, mg/dL	160 (139–186)	166 (143–193)
HDL-cholesterol, mg/dL	49 (38–59)	49 (38–61)
CD4	386 (258–660)	624 (407–720)
HIV viral load, log	0 (0–0)	0 (0–0)

Data expressed as median (interquartile range) or absolute (%). ALT: alanine transaminase; AST: aspartate transaminase; GGT: gamma-glutamyltransferase; HDL: high-density lipoprotein; HCV: hepatitis C virus; HIV: human imunodeficiency virus; kPa: kiloPascal. Missing data (n) for the intra/interobserver analysis: HCV genotype (73/42); HCV treatment (83/56); ALT (56/33); AST (57/32); AP (60/37); GGT (58/33); total bilirubin (71/38); triglycerides (76/79); total cholesterol (75/78); HDL-cholesterol (77/85). CD4 and HIV viral load was measured only in HIV co-infected patients.

elastography. Our findings on reproducibility of transient elastography were similar to previously reported and fill the gap of knowledge concerning the learning curve of transient elastography.

Initially, Fraquelli et al. reported excellent overall intra- and interobserver agreement (ICC of 0.98) in 200 patients with chronic liver disease, most of them with chronic hepatitis C [5]. Although excellent rates of intraobserver (ICC of 0.94) and interobserver agreement (ICC of 0.93), Boursier et al. reported 25% of discrepancy of at least one stage of fibrosis between two experimented operators [6]. These non-negligible discordance rates were confirmed by others authors in a study with HIV patients [7]. More recently, Nascimbeni et al. reported a discrepancy of 32% for one-fibrosis stage and 10% for two or more stages of fibrosis in paired liver stiffness measurements performed in patients with chronic liver diseases [13]. Furthermore, a previous study from our team analyzed the interobserver variability of transient elastography in a homogeneous population of patients with chronic hepatitis C recruited in a different center in Rio de Janeiro, Brazil [14]. We reported up to 35% of discordance of at least one stage of fibrosis although a high correlation between measures performed by two experimented operators. In a population of 188 patients with HIV/HCV co-infection, Neukam et al. reported high rates of interobserver agreement [ICC = 0.976 (95% CI 0.968–0.982] and k values of 0.60 and 0.89 for diagnosis of fibrosis stages $F \geq 2$ and cirrhosis, respectively [15]. The present study described very similar results on overall agreement [intraobserver ICC of 0.935 (95% CI 0.891–0.979) and interobserver ICC of 0.955 (95% CI 0.932–0.978)] and k values ranging from 0.69 to 0.85 for stage liver fibrosis in patients with HCV/HIV co-infection. We also reported around 20% of discrepancy for at least one stage of fibrosis and 5% for two or more stages of fibrosis between two operators or a single operator performing repeated measures in patients with HCV/HIV (supplementary material, Table S1).

We acknowledge that it is very difficult, even to liver biopsy (considered as the gold standard), to well discriminate adjacent stages of fibrosis, especially intermediate stages [16]. However, these high rates of discrepancy of at least one stage of fibrosis are not negligible and might influence the management of patients with chronic hepatitis C. This variability might be associated with the operator's experience but also related others factors, such as the type of probe, patient's fasting status, respiratory phase of measurement acquisition and probe position. Liver stiffness measurement performed with the M probe has been described as higher (mean of 1.5 kPa) compared to those with the XL probe [17]. In addition, liver stiffness increases by food intake [18], presence of necro-inflammatory activity [19], extrahepatic cholestasis [20] or liver congestion [21].

Table 2 Intra- and interobserver agreement in transient elastography according to the training operator experience.

	Intraobserver agreement				
Training operator experience (exams)	0–100	101–300	301–500	> 500	Overall
	$n = 13$	$n = 38$	$n = 31$	$n = 43$	$n = 125$
Kappa (SE)					
For any stage of fibrosis	0.54 (0.17)	0.72 (0.11)	0.54 (0.11)	0.66 (0.09)	0.67 (0.05)
For fibrosis stages $F \geq 2$	0.56 (0.25)	0.78 (0.16)	0.67 (0.18)	0.46 (0.13)	0.74 (0.09)
For fibrosis stages $F \geq 3$	0.58 (0.25)	1.00 (0.16)	0.78 (0.18)	0.72 (0.15)	0.84 (0.09)
For fibrosis stages $F = 4$	0.63 (0.26)	0.84 (0.16)	0.61 (0.18)	0.81 (0.15)	0.77 (0.09)
Correlation between examinations					
First LSM (IQR)[a]	7.6 (6.7–9.8)	6.5 (5.4–12.1)	12.0 (8.6–20.4)	11.6 (9.4–19.0)	10.0 (6.8–15.1)
Second LSM (IQR)[a]	6.6 (4.9–9.1)	6.5 (4.8–11.1)	10.5 (8.6–19.2)	11.3 (8.1–18.0)	9.5 (6.3–14.9)
First vs. second LSM P-value[b]	0.030	0.663	0.070	0.388	0.015
Spearman's rho (P-value)	0.586 (0.0353)	0.872 (< 0.0001)	0.820 (< 0.0001)	0.888 (< 0.0001)	0.906 (< 0.0001)
Intraclass correlation coefficient (95% CI)	0.748 (0.504–0.993)	0.966 (0.945–0.987)	0.838 (0.732–0.944)	0.963 (0.942–0.965)	0.926 (0.901–0.951)
Discrepancy on staging liver fibrosis, n (%)					
At least one stage	4 (31)	6 (16)	10 (32)	10 (23)	30 (24)
Two or more stages	2 (15)	0 (0)	2 (7)	3 (7)	7 (6)

	Interobserver agreement			
Training operator experience (exams)	0–100	101–300	301–500	Overall
	$n = 45$	$n = 60$	$n = 46$	$n = 151$
Kappa (SE)				
For any stage of fibrosis	0.74 (0.09)	0.57 (0.08)	0.66 (0.09)	0.65 (0.05)
For fibrosis stages $F \geq 2$	0.86 (0.15)	0.83 (0.13)	0.87 (0.15)	0.85 (0.08)
For fibrosis stages $F \geq 3$	0.77 (0.15)	0.72 (0.13)	0.73 (0.15)	0.74 (0.08)
For fibrosis stages $F = 4$	0.85 (0.15)	0.71 (0.13)	0.69 (0.15)	0.75 (0.08)
Correlation between examinations				
T-operator LSM (IQR)[a]	9.1 (6.1–16.5)	8.1 (6.1–12.7)	7.5 (4.9–11.9)	8.1 (5.6–13.0)
E-operator LSM (IQR)[a]	8.6 (5.9–16.1)	8.7 (5.9–14.1)	8.1 (5.3–11.9)	8.4 (5.7–13.8)
T vs. E-operator LSM P-value[b]	0.79	0.453	0.523	0.474
Spearman's rho (P-value)	0.917 (< 0.0001)	0.899 (< 0.0001)	0.893 (< 0.0001)	0.907 (< 0.0001)
Intraclass correlation coefficient (95% CI)	0.953 (0.926–0.980)	0.963 (0.944–0.981)	0.721 (0.581–0.860)	0.912 (0.885–0.939)
Discrepancy on staging liver fibrosis, n (%)				
At least one stage	8 (18)	18 (30)	11 (24)	37 (25)
Two or more stages	2 (4)	2 (3)	2 (4)	6 (4)

CI: confidence interval; E: experimented; IQR: interquartile range; LSM: liver stiffness measurement; SE: standard error; T: training.

[a] LSM expressed as median (IQR).

[b] Comparison by Mann-Whitney test.

Table 3 Intra- and interobserver agreement in transient elastography using a 100-exam and a 300-exam experience of the training operator for definition of an experimented operator.

	Intraobserver agreement (n = 125)		Interobserver agreement (n = 151)	
Definition of experimented operator using a cut-off = 100 exams				
Training operator experience	≤ 100	> 100	≤ 100	> 100
	n = 13	n = 112	n = 45	n = 106
Kappa (SE)				
For any stage of fibrosis	0.54 (0.17)	0.68 (0.06)	0.74 (0.09)	0.61 (0.06)
For fibrosis stages F ≥ 2	0.56 (0.25)	0.76 (0.09)	0.86 (0.15)	0.85 (0.10)
For fibrosis stages F ≥ 3	0.58 (0.25)	0.86 (0.09)	0.77 (0.15)	0.73 (0.10)
For fibrosis stages F = 4	0.63 (0.26)	0.77 (0.09)	0.85 (0.15)	0.70 (0.10)
Correlation between examinations				
First LSM (IQR)[a]	7.6 (6.7–9.8)	10.8 (6.8–15.5)	9.1 (6.1–16.5)	7.8 (5.4–12.6)
Second LSM (IQR)[a]	6.6 (4.9–9.1)	10.0 (6.4–16.5)	8.6 (5.9–16.1)	8.4 (5.7–12.8)
First vs. second LSM P-value[b]	0.030	0.070	0.790	0.308
Spearman's rho (P-value)	0.586 (0.0353)	0.921 (< 0.0001)	0.917 (< 0.0001)	0.904 (< 0.0001)
ICC (95% CI)	0.748 (0.504–0.993)	0.930 (0.905–0.955)	0.953 (0.926–0.980)	0.894 (0.855–0.932)
Discrepancy on staging liver fibrosis, n (%)				
At least one stage	4 (31)	26 (23)	8 (18)	29 (27)
Two or more stages	2 (15)	5 (5)	2 (4)	4 (4)
Definition of experimented operator using a cut-off = 300 exams				
Training operator experience	≤ 300	> 300	≤ 300	> 300
	n = 51	n = 74	n = 105	n = 46
Kappa (SE)				
For any stage of fibrosis	0.68 (0.09)	0.61 (0.07)	0.64 (0.06)	0.66 (0.09)
For fibrosis stages F ≥ 2	0.72 (0.14)	0.58 (0.11)	0.84 (0.10)	0.87 (0.15)
For fibrosis stages F ≥ 3	0.90 (0.14)	0.75 (0.12)	0.74 (0.10)	0.73 (0.15)
For fibrosis stages F = 4	0.81 (0.14)	0.73 (0.12)	0.77 (0.10)	0.69 (0.15)
Correlation between examinations				
First LSM (IQR)[a]	6.8 (5.7–11.8)	11.8 (8.8–19.0)	8.7 (6.1–14.0)	7.5 (4.9–11.9)
Second LSM (IQR)[a]	6.6 (4.8–10.6)	11.2 (8.5–18.2)	8.6 (5.9–14.4)	8.1 (5.3–11.9)
First vs. second LSM P-value[b]	0.098	0.064	0.654	0.523
Spearman's rho (P-value)	0.809 (< 0.0001)	0.871 (< 0.0001)	0.905 (< 0.0001)	0.893 (< 0.0001)
ICC (95% CI)	0.933 (0.897–0.967)	0.915 (0.878–0.952)	0.959 (0.943–0.974)	0.721 (0.581–0.860)
Discrepancy on staging liver fibrosis, n (%)				
At least one stage	10 (20)	20 (27)	26 (25)	11 (24)
Two or more stages	2 (4)	5 (7)	4 (4)	2 (4)

CI: confidence interval; ICC: intraclass correlation coefficient; IQR: interquartile range; LSM: liver stiffness measurement; SE: standard error. The first LSM was performed by the training operator in interobserver analysis.

[a] LSM expressed as median (IQR).

[b] Comparison by Mann-Whitney test.

Furthermore liver stiffness measurement might be influenced by probe position [22] or operator experience [10,13]. In the present study, repeated measures intra- or interobserver were performed in the same patient's fasting status, acute hepatitis cases were excluded, patients had normal total bilirubin and no signs of heart failure. In addition, the M probe should be placed perpendicularly on the reference point (i.e., intersection between the xyphoid line and the median axillary line). However, measurement spots were chosen by the respective operator during examination and the intercostal space or respiratory phase of each measure was not recorded. Several factors other than operator experience could impact the liver stiffness measurement. A discrepancy of one liver fibrosis stage might be related to the transient elastography technique itself and few factors that cannot be controlled during examination. Thus, the real disagreement between liver stiffness measurement and others noninvasive methods, such as biomarkers, should be considered in the presence of a discrepancy of two or more fibrosis stages. However, we suggest that noninvasive

ARTICLE IN PRESS

H. Perazzo et al.

Table 4 Overall intra- and interobserver agreement using cut-offs of transient elastography proposed by Ziol et al. [23] and Fraquelli et al. [5] for all patients and combining cut-offs proposed by Castera et al. [10] for HCV mono-infected and Vergara et al. [24] for HCV/HIV co-infected patients.

Classification system	Intraobserver analysis ($n = 125$)			Interobserver analysis ($n = 151$)		
	Ziol et al., for all	Fraquelli et al., for all	Vergara et al., for HCV/HIV Castera et al., for HCV	Ziol et al., for all	Fraquelli et al., for all	Vergara et al., for HCV/HIV Castera et al., for HCV
Kappa (SE)						
For any stage of fibrosis	0.79 (0.06)	0.66 (0.06)	0.65 (0.05)	0.69 (0.05)	0.66 (0.05)	0.68 (0.05)
For fibrosis stages F ≥ 2	0.85 (0.09)	0.80 (0.09)	0.74 (0.09)	0.82 (0.08)	0.85 (0.08)	0.84 (0.08)
For fibrosis stages F ≥ 3	0.84 (0.09)	0.76 (0.09)	0.82 (0.09)	0.74 (0.08)	0.70 (0.08)	0.78 (0.08)
For fibrosis stages F = 4	0.77 (0.09)	0.76 (0.09)	0.78 (0.09)	0.75 (0.08)	0.74 (0.08)	0.73 (0.08)
Discrepancy on staging liver fibrosis, n (%)						
At least one stage	18 (14)	32 (26)	32 (26)	30 (20)	34 (23)	34 (23)
Two or more stages	7 (6)	9 (7)	5 (4)	10 (7)	11 (7)	6 (4)

HCV: hepatitis C virus; HIV: human immunodeficiency virus; SE: standard error. Classification proposed by Ziol et al.: < 8.5 as F0F1; 8.5—9.4 as F2; 9.5—14.4 as F3 and > 14.4 kPa as F4. Classification proposed by Fraquelli et al.: < 7.9 as F0F1; 7.9—10.3 as F2; 10.4—11.9 as F3 and > 11.9 kPa as F4. Classification proposed by Castera et al.: < 7.1 as F0F1; 7.1—9.4 as F2; 9.5—12.4 as F3 and > 12.4 kPa as F4. Classification proposed by Vergara et al. for co-infected HCV/HIV patients: < 7.2 as F0F1; 7.2—11.8 as F2; 11.9—14.5 as F3 and > 14.5 kPa as F4.

evaluation of liver fibrosis, including transient elastography results, should be carefully analyzed by expert hands.

Blood tests can accurately stage liver fibrosis in patients with chronic hepatitis C [23]. FibroTest®-ActiTest® was validated for definition of inactive carrier of chronic hepatitis B and transient elastography or liver biomarkers can be used to rank severity of liver disease from no-fibrosis to death in patients with chronic viral hepatitis [24,25]. However, transient elastography had a lower applicability compared to FibroTest® (80% vs. 98%) [8,26] and a relative high rate of variability of liver stiffness measurement [7,14].

Regarding the excellent overall correlation between examinations, the interpretation of artificial and non-linear cut-off distribution should explain this relative high discordance for staging liver fibrosis by transient elastography. In the present study used the most validated classification system available to convert liver stiffness in METAVIR fibrosis stages for patients with chronic hepatitis C [12]. However, different cut-offs could be used in patients with HIV co-infection. In a sensitivity analysis, we reported similar overall agreement rates when repeating the analysis using two others classifications systems for chronic hepatitis C [5,27] and also using different cut-offs according to presence of HIV co-infection [28] (Table 4). Transient elastography performance might also be influenced by the presence of HIV co-infection [29] and the prevalence of liver fibrosis stages, known as spectrum effect [30]. In our study, the overall agreement was similar between patients with and without HIV infection (supplementary material, Table S1). However, proportion of HIV infected patients was different according to training operator experience in both analysis and liver stiffness was significantly higher when the training operator's experience was at least high in patients included in the intraobserver analysis (supplementary material, Table S2). In addition, patients included in both cohorts for evaluation of intra- and interobserver agreement had a major difference in the prevalence of mild fibrosis (F0F1): 28% vs. 43% for intra- and interobserver analysis, respectively. This fact might explain the lack of improvement on agreement rates related to training operator experience in the interobsever analysis.

The present study confirmed that transient elastography might be influenced by an operator effect [10,13]. First studies have reported a 50-exams cut-off for definition of an experimented operator [31] followed by an increase of this cut-off up to 500-exam experience [8]. However, no study has evaluated a complete training program following an individual from initially inexperienced to experimented operator according to the current definition. Intra- and interobserver agreements were satisfactory when the training operator experience was at least moderate and did not improve following the increase of number of examinations beyond this cut-off. Thus, our findings suggest that performance of 100 examinations could be implemented to define an experimented operator. In continental countries with a high prevalence of chronic liver disease, such as Brazil, a more rapid training of operators is essential to permit the assessment of this new technology in large scale to the majority of population. Nowadays, transient elastography is available only in tertiary centers due to be an expensive device but also due to the lack of experimented operators because the need of a long and intensive training program.

The main limitations of our study were its retrospective design, the fact that training and experimented operators were not blinded to his/her colleague results and the lack of liver biopsy for correctly staging liver fibrosis to be taken as an independent control for fibrosis concordance assessment. The aim of the training program was to form the training operator to correctly assess liver fibrosis by transient elastography. Thus, it was not feasible to blind the training operator to his first examination and the experimented operator for the performance of the examination by the training operator. The lack of liver biopsies as a reference for staging fibrosis can be explained by the fact that currently, for clinical practice, liver fibrosis has been accurately assessed by transient elastography and liver biopsy should be restricted to few patients. Other limitations should be the absence of assessment of the impact of hepatic necro-inflammatory activity or steatosis on liver stiffness.

Our major strength was the assessment of both intra- and interobserver variability of transient elastography in a representative sample of patients with chronic hepatitis C with or without HIV co-infection. The originality of the study rests on the evaluation of the learning curve of transient elastography. The present study followed the complete training of an initially inexperienced operator on transient elastography over almost 2 years until the performance of at least 500 examinations under supervision.

In conclusion, although there was a considerable discrepancy on fibrosis staging between examinations and a small power, transient elastography had an acceptable reproducibility in our population. Performance of at least 100 examinations should be used to define an experimented operator.

Authors' involvements

Hugo Perazzo: study concept and design, statistical analysis, interpretation of data, manuscript drafting and critical revision; Flavia F. Fernandes: performance of exams, analysis and interpretation of data, critical revision of the manuscript; João Carlos Soares: performance of exams, critical revision of the manuscript; Juliana Fittipaldi: data collection; Sandra W. Cardoso: analysis and interpretation of data, critical revision of the manuscript; Beatriz Grinsztejn: study concept and design, analysis and interpretation of data, critical revision of the manuscript; Valdiléa G. Veloso: study concept and design, analysis and interpretation of data, critical revision of the manuscript, study supervision. All authors approved the final version of the manuscript.

Disclosure of interest

The authors declare that they have no conflicts of interest concerning this article.

Acknowledgements

Appendix A. Supplementary material

Supplementary material associated with this article can be found in the online version available at http://dx.doi.org/10.1016/j.clinre.2015.04.004.

References

[1] Mohd Hanafiah K, Groeger J, Flaxman AD, Wiersma ST. Global epidemiology of hepatitis C virus infection: new estimates of age-specific antibody to HCV seroprevalence. Hepatology 2013;57:1333–42.

[2] Lacombe K, Rockstroh J. HIV and viral hepatitis co-infections: advances and challenges. Gut 2012;61:47–58.

[3] Mutimer D, Aghemo A, Diepolder H, Negro F, Robaeys G, Ryder S, et al. EASL Clinical Practice Guidelines: management of hepatitis C virus infection. J Hepatol 2014;60:392–420.

[4] Bota S, Herkner H, Sporea I, Salzl P, Sirli R, Neghina AM, et al. Meta-analysis: ARFI elastography versus transient elastography for the evaluation of liver fibrosis. Liver Int 2013;33:1138–47.

[5] Fraquelli M, Rigamonti C, Casazza G, Conte D, Donato MF, Ronchi G, et al. Reproducibility of transient elastography in the evaluation of liver fibrosis in patients with chronic liver disease. Gut 2007;56:968–73.

[6] Boursier J, Konate A, Gorea G, Reaud S, Quemener E, Oberti F, et al. Reproducibility of liver stiffness measurement by ultrasonographic elastometry. Clin Gastroenterol Hepatol 2008;6:1263–9.

[7] Roca B, Resino E, Torres V, Herrero E, Penades M. Interobserver discrepancy in liver fibrosis using transient elastography. J Viral Hepat 2012;19:711–5.

[8] Castera L, Foucher J, Bernard PH, Carvalho F, Allaix D, Merrouche W, et al. Pitfalls of liver stiffness measurement: a 5-year prospective study of 13,369 examinations. Hepatology 2010;51:828–35.

[9] de Ledinghen V, Vergniol J. Transient elastography (FibroScan®). Gastroenterol Clin Biol 2008;32:58–67.

[10] Poynard T, Ingiliz P, Elkrief L, Munteanu M, Lebray P, Morra R, et al. Concordance in a world without a gold standard: a new noninvasive methodology for improving accuracy of fibrosis markers. PLoS One 2008;3:e3857.

[11] Intraobserver and interobserver variations in liver biopsy interpretation in patients with chronic hepatitis C. The French METAVIR Cooperative Study Group. Hepatology 1994;20:15–20.

[12] Castera L, Vergniol J, Foucher J, Le Bail B, Chanteloup E, Haaser M, et al. Prospective comparison of transient elastography, FibroTest®, APRI, and liver biopsy for the assessment of fibrosis in chronic hepatitis C. Gastroenterology 2005;128:343–50.

[13] Nascimbeni F, Lebray P, Fedchuk L, Oliveira CP, Alvares-da-Silva MR, Varault A, et al. Significant variations in elastometry measurements made within short-term in patients with chronic liver diseases. Clin Gastroenterol Hepatol 2015;13(4):763.e1-6–71.e1-6, http://dx.doi.org/10.1016/j.cgh.2014.07.037.

[14] Perazzo H, Fernandes FF, Gomes A, Terra C, Perez RM, Figueiredo FA. Interobserver variability in transient elastography analysis of patients with chronic hepatitis C. Liver Int 2015;35(5):1533–9, http://dx.doi.org/10.1111/liv.12551.

[15] Neukam K, Recio E, Camacho A, Macias J, Rivero A, Mira JA, et al. Interobserver concordance in the assessment of liver fibrosis in HIV/HCV co-infected patients using transient elastometry. Eur J Gastroenterol Hepatol 2010;22:801–7.

[16] Poynard T, Lenaour G, Vaillant JC, Capron F, Munteanu M, Eyraud D, et al. Liver biopsy analysis has a low level of performance for diagnosis of intermediate stages of fibrosis. Clin Gastroenterol Hepatol 2012;10:657–63.

[17] de Ledinghen V, Wong VW, Vergniol J, Wong GL, Foucher J, Chu SH, et al. Diagnosis of liver fibrosis and cirrhosis using liver stiffness measurement: comparison between M and XL probe of FibroScan®. J Hepatol 2012;56:833–9.

[18] Arena U, Lupsor Platon M, Stasi C, Moscarella S, Assarat A, Bedogni G, et al. Liver stiffness is influenced by a standardized meal in patients with chronic hepatitis C virus at different stages of fibrotic evolution. Hepatology 2013;58:65–72.

[19] Fung J, Lai CL, Cheng C, Wu R, Wong DK, Yuen MF. Mild-to-moderate elevation of alanine aminotransferase increases liver stiffness measurement by transient elastography in patients with chronic hepatitis B. Am J Gastroenterol 2011;106:492–6.

[20] Millonig G, Reimann FM, Friedrich S, Fonouni H, Mehrabi A, Buchler MW, et al. Extrahepatic cholestasis increases liver stiffness (FibroScan®) irrespective of fibrosis. Hepatology 2008;48:1718–23.

[21] Colli A, Pozzoni P, Berzuini A, Gerosa A, Canovi C, Molteni EE, et al. Decompensated chronic heart failure: increased liver stiffness measured by means of transient elastography. Radiology 2010;257:872–8.

[22] Ingiliz P, Chhay KP, Munteanu M, Lebray P, Ngo Y, Roulot D, et al. Applicability and variability of liver stiffness measurements according to probe position. World J Gastroenterol 2009;15:3398–404.

[23] Chou R, Wasson N. Blood tests to diagnose fibrosis or cirrhosis in patients with chronic hepatitis C virus infection. Ann Intern Med 2013;158:807–20.

[24] Poynard T, Vergniol J, Ngo Y, Foucher J, Thibault V, Munteanu M, et al. Staging chronic hepatitis B into seven categories, defining inactive carriers and assessing treatment impact using a fibrosis biomarker (FibroTest®) and elastography (FibroScan®). J Hepatol 2014;61:994–1003.

[25] Poynard T, Vergniol J, Ngo Y, Foucher J, Munteanu M, Merrouche W, et al. Staging chronic hepatitis C in seven categories using fibrosis biomarker (FibroTest®) and transient elastography (FibroScan®). J Hepatol 2014;60:706–14.

[26] Poynard T, Munteanu M, Deckmyn O, Ngo Y, Drane F, Messous D, et al. Applicability and precautions of use of liver injury biomarker FibroTest®. A reappraisal at 7 years of age. BMC Gastroenterol 2011;11:39.

[27] Ziol M, Handra-Luca A, Kettaneh A, Christidis C, Mal F, Kazemi F, et al. Noninvasive assessment of liver fibrosis by measurement of stiffness in patients with chronic hepatitis C. Hepatology 2005;41:48–54.

[28] Vergara S, Macias J, Rivero A, Gutierrez-Valencia A, Gonzalez-Serrano M, Merino D, et al. The use of transient elastometry for assessing liver fibrosis in patients with HIV and hepatitis C virus co-infection. Clin Infect Dis 2007;45:969–74.

[29] Frulio N, Trillaud H, Perez P, Asselineau J, Vandenhende M, Hessamfar M, et al. Acoustic radiation force impulse (ARFI) and transient elastography (TE) for evaluation of liver fibrosis in HIV-HCV co-infected patients. BMC Infect Dis 2014;14:405.

[30] Poynard T, de Ledinghen V, Zarski JP, Stanciu C, Munteanu M, Vergniol J, et al. FibroTest® and FibroScan® performances revisited in patients with chronic hepatitis C. Impact of the spectrum effect and the applicability rate. Clin Res Hepatol Gastroenterol 2011;35:720–30.

[31] Boursier J, Konate A, Guilluy M, Gorea G, Sawadogo A, Quemener E, et al. Learning curve and interobserver reproducibility evaluation of liver stiffness measurement by transient elastography. Eur J Gastroenterol Hepatol 2008;20:693–701.