Point quantification elastography in the evaluation of liver elasticity in healthy volunteers: a reliability study based on operator expertise

Cristina Felicani¹ · Chiara De Molo¹ · Horia Stefanescu^{2,3} · Fabio Conti³ · Elena Mazzotta¹ · Veronica Gabusi¹ · Elena Nardi³ · Antonio Maria Morselli-Labate³ · Pietro Andreone³ · Carla Serra¹

Received: 4 February 2018 / Accepted: 17 April 2018 / Published online: 22 May 2018
© Società Italiana di Ultrasonologia in Medicina e Biologia (SIUMB) 2018

Abstract

Purpose The assessment of liver fibrosis is essential in the management of patients with chronic liver diseases. Liver biopsy is considered the gold standard procedure for this purpose, though the recent development of new elastosonographic techniques to measure liver stiffness (LS) noninvasively is promising. Point quantification elastography (PQE) showed good results but less is known about the level of skill needed to obtain reliable results. The aim of the study was to evaluate the reproducibility of PQE in assessing LS in healthy subjects comparing three operators with different expertise.

Methods Between December 2012 and April 2013, 50 consecutive healthy volunteers (18 males, 32 females), median age 30 years (range 25–66) and BMI 22.4 (range 16.7–33.6) were submitted to PQE (iU22 Philips, Bothell, WA, USA) by three operators: two US and elastography providers (one expert and one with intermediate skill) and a skilled transient elastography (Fibroscan) operator with no expertise in US. Intra- and inter-observer agreements were assessed by intraclass correlation coefficient (ICC).

Results PQE measurement was obtained in all subjects by all evaluators. No significant differences of mean liver stiffness were found among operators ($P=0.980$). Intra-observer agreement was excellent 0.918 (0.941 for expert, 0.917 for intermediate and 0.888 for novice). The ICC of the inter-observer agreement among the three raters was excellent (0.882) and was higher in normal than overweight patients (0.923 vs. 0.603; $P=0.011$).

Conclusion PQE is a reliable and reproducible non-invasive method for the assessment of LE, and can be performed also by a non-experienced operator.

Keywords Elastosonography · Reproducibility · Liver stiffness · Elastography · Point shear wave

Sommario

Scopo La valutazione della fibrosi epatica è fondamentale nella gestione dei pazienti con epatopatia cronica. A tutt'oggi il gold standard diagnostico resta la biopsia epatica anche se ha trovato sempre più spazio l'utilizzo delle tecniche elastografiche per misurare la stiffness epatica (LS). L'elastografia *Point Quantification* (PQE) ha mostrato buoni risultati in questo campo, anche se non è chiaro quale sia il livello di esperienza ecografica necessario perché l'operatore possa ottenere risultati riproducibili. Lo scopo dello studio è stato quello di valutare la riproducibilità della metodica analizzando le misurazioni di LS effettuate, in soggetti sani, da 3 operatori con diversi livelli di esperienza in ambito ecografico.

Metodi Sono stati sottoposti PQE (iU22 Philips, Bothell, WA, USA) 50 volontari sani (18M, 32 F), età media 30 anni (range 25–66) e BMI 22.4 (range 16.7–33.6). Per ciascun soggetto sono state ottenute misurazioni di LS da 3 differenti operatori, due con esperienza sia ecografica che elastografica (uno esperto ed uno con esperienza intermedia) ed un operatore con esperienza in elastografia Transient (Fibroscan), ma senza un background ecografico (novizio). La concordanza intra ed inter osservatore sono state valutate con il coefficiente di correlazione interclasse (ICC).

Risultati Non sono emerse differenze significative nella misurazione della LS media fra i 3 operatori ($P=0.980$). La concordanza intra-operatore è risultata eccellente 0.918 (0.941 per l'operatore esperto, 0.917 per l'intermedio and 0.888 per il novizio). Anche la concordanza inter-operatore è risultata eccellente (0.882) e superiore nei pazienti normopeso rispetto ai pazienti sovrappeso (0.923 vs. 0.603; $P=0.011$).

Conclusioni PQE è un sistema attendibile, riproducibile, non invasivo per la valutazione della LS e può essere eseguito anche da operatori senza esperienza ecografica.

Introduction

Liver fibrosis quantification plays a crucial role in the management of patients with chronic liver disease (CLD) as a prognostic hallmark for the detection, surveillance and staging of the disease and its complications, and as a guide for the management of antiviral treatment. Nowadays, liver biopsy is still considered the gold standard to assess the stage of fibrosis [1], although an imperfect one, mainly because of sampling variability issues and some limitations, including its invasiveness, with severe potential complications limiting repeated assessments [2]. Furthermore, sampling variability can affect the histological evaluation: the optimal liver biopsy specimen should be 20 mm in length with 11 portal tracts, but it is not often obtained in clinical practice [3]; and this is an important limitation considering that the sample represents only 1/50,000 of liver tissue [3].

In the last ten years, several new diagnostic methods, and in particular, elastosonographic techniques have been intensively evaluated for their ability to estimate the amount of liver fibrosis in a non-invasive, accurate, non expensive, reliable and reproducible manner.

Ultrasound elastography technologies are all based on the measurement of deformation induced in tissue by a static/quasi-static or dynamic stress. European federation of Societies for Ultrasound in Medicine and Biology (EFSUMB) and World Federation for Ultrasound in Medicine and Biology (WFUMB) Guidelines summarize the features and clinical application of all these techniques [4–9].

Transient elastography (TE, FibroScan®, Echosens, Paris, France) has been first proposed for the assessment of liver stiffness (LS) as an indirect index of hepatic fibrosis and is nowadays widely applied in daily clinical practice to estimate liver fibrosis. The diagnostic performance of TE in patients with HCV-related liver disease has been evaluated in three important multicenter studies [9–11], and it has also been tested in patients with NAFLD [12]. Intra- and inter-observer reproducibility of the technique are good [13]; the disadvantages of TE assessment of fibrosis are the impossibility to discriminate between the various contiguous degrees of fibrosis [14], false overestimated results during ALT flares [15], the high cost of the dedicated device, and measurement failure in patients with ascites and poor acoustic windows [5–10]. Furthermore, a training course is necessary to instruct the performers [12].

Point quantification elastography (PQE) with ElastPQ® (Philips Healthcare, Bothell, WA, USA) is a one-dimensional

technique integrated onto a conventional ultrasound system and the examination can be performed with the same device used for conventional US scans. Unlike transient elastography, PQE relies on the mechanical excitation of the tissue by the shear waves produced by a localized, impulsive, acoustic radiation force and it is classified as an ARFI-based technology. An interrogation region (ROI—region of interest) can be placed anywhere in the US visualized liver parenchyma. Immediately after the image guided acquisition, the machine displays the result (in kPa) and mean, median values and the deviations of valid measurement. Elastic value E (in kPa) is calculated using the equation $E = 3 \rho c_s^2$ where c_s (in m/s) is defined as the shear wave propagation velocity and ρ is the tissue density (whose approximated value in the human body is 1). According to this formula, stiffer is the tissue then faster the shear wave propagates [16–18].

The few data available in the literature for this technique show a good reproducibility [16], but not much is known about the necessity of training, the level of expertise needed to perform a trustworthy examination and at present there is no agreement as to what constitutes an experienced operator for PQE [8].

The aim of this study was to evaluate the reproducibility of PQE in assessing normal liver elasticity among operators with different levels of expertise, and to investigate any training effect.

Materials and methods

Study design

This study was a single-center prospective study. Healthy subjects were enrolled to evaluate liver PQE measurements by three raters, each of them performing LS measurements in the same subject in the same day. Two raters were experienced in ultrasound and elastography: one with more than 5 years of expertise (US expert: C.S.) and one with 2 years of expertise (US intermediate: C.D.M.), while the third had no expertise in US, but 5 years of expertise in transient elastography (novice: F.C.) and he received a brief explanation on how to perform the measurements. All raters were blinded about the results obtained by others.

Subjects

Fifty volunteers, drawn from a pool of hospital staff members, were evaluated from December 2012 to April 2013. All

of them had normal liver parameters and US without known liver diseases. All subjects underwent a preliminary US scan to confirm the absence of hepatic parenchymal abnormalities (i.e., liver lesions, steatosis, etc.). The age ranged between 25 and 66 years with a mean of 34.2 ± 10.4 years (median: 30 years; IQR range: 28–36 years) with a male-to-female ratio of 0.56 (18 males, 76.0%; 32 females, 24.0%). The body mass index (BMI) ranged between 16.7 and 33.6 kg/m^2 with a mean of $22.8 \pm 3.2 \text{ kg/m}^2$ (median: 22.4 kg/m^2 ; IQR range: $20.8\text{--}24.9 \text{ kg/m}^2$). Twelve out of 50 subjects (24.0%) had a BMI greater than 25 kg/m^2 ; the BMI of these subjects ranged between 25.2 and 33.6 kg/m^2 with a mean value of 27.2 kg/m^2 (median: 26.5 kg/m^2 ; IQR range: $25.6\text{--}28.1 \text{ kg/m}^2$) while the BMI of the 38 subjects with normal BMI had the following descriptive statistics: mean: 21.4 kg/m^2 ; SD: 1.9 kg/m^2 ; range: $16.7\text{--}24.8 \text{ kg/m}^2$; median: 21.9 kg/m^2 ; IQR range: $20.2\text{--}22.6 \text{ kg/m}^2$.

Ultrasound imaging protocol

Ultrasound evaluation was performed using an iU22 ultrasound machine (Philips, Bothell, WA, USA, software version 6.3.2.2) with the C5-1 curvilinear probe. Using B-Mode US image as guidance a ROI of $0.5 \times 1.5 \text{ cm}$ was freely located in liver parenchyma at 3–8 cm from the capsule, avoiding vascular structures or artifacts. When the measurement returned a value equal to 0, often caused by an excessive target motion or poor beam coupling, acquisitions were identified as not quantifiable [4–9].

The mean of ten valid acquisitions was considered representative for LS measurement as recommended [5, 8, 19].

To diminish potential perturbing factors, the subjects were examined in fasting conditions, in supine position with right arm in maximal abduction and were asked to suspend their breath in intermediate expiration for 5–10 s during each measurement, since these situations may influence LS in healthy subjects [5–8].

All LS measurements have been obtained from the right lobe with an intercostal approach [5, 17].

Main outcome measurements

The intra-operator variability was evaluated by assessing the agreement among all the ten measurements made by the same operator.

The inter-operator variability was assessed by comparing the mean values of the ten measurements obtained in the same day by each operator among the three operators, as well as between pairs of operators.

To establish if a training effect was present, both the intra- and inter-operator agreements observed in the last 25 subjects were compared with the agreements found in the first 25 subjects.

In addition, we have evaluated if a reduction of the number of valid shear wave acquisitions could provide reliable results. For this purpose, we evaluated the inter- and intra-operator agreements using two subsets of five measurements (i.e., the first five and last five acquisitions) and we have compared them with the agreement obtained by the entire set of ten measurements.

Statistical analysis

Mean $\pm$ standard deviation (SD), range, median, interquartile range (IQR) and absolute and relative frequencies were used as descriptive statistics. Non-parametric statistics were used to analyze stiffness values: the Friedman's two-way ANOVA was applied to comparisons among the three observers and the Kruskal–Wallis test was applied to comparisons between different groups of patients.

Reproducibility and reliability were assessed through the calculation of the average measure intraclass correlation coefficient (ICC) [19]. The ICCs were reported together with their 95% confidence intervals (CIs). ICC ranges in values from 0 to +1; and in values of ICC below 0.40 or so may be taken to represent poor reliability, values above 0.75 to represent excellent reliability, and values between 0.40 and 0.75 to represent fair to good reliability [20]. The ICCs among the ten measurements made by each operator in the 50 patients were taken into account to assess the intra-operator agreement and the overall agreement was evaluated by means of the same ICC computed by pooling together the data of the three operators. The inter-observer variability was evaluated by estimating the LS of each patient by means of the average of the ten measurements made by each operator. The overall ICC among the three operators was evaluated as well as the ICCs of the three pairs of operators. The differences between the measurements made by pairs of operators were computed and plotted (ordinate value) against the mean of the two measurements as the abscissa value (expected value) according to the method described by Bland and Altman [21]. This analysis was able to: (i) assess the existence of any systematic difference between the measurements (i.e., fixed bias); (ii) investigate any possible relationship of the discrepancies between the measurements and the expected value (i.e., proportional bias); and (iii) identify the presence of outliers. The exact lower and the upper limits of agreement (together with their 95% CIs) were also computed [22] and the repeatability coefficients were evaluated according to the British Standards Institution [23]. The proportional bias was tested by means of the Pearson's r coefficient of correlation between differences and average values.

To compare the reproducibility pairs of ICC coefficients, the z -test was applied to the corrected z -transformations of the ICC coefficients evaluated according to Fisher [24]. The standard errors of the corrected z -transformations of the ICC

coefficients were estimated as the 3.92 (i.e., 1.96×2) fraction of the differences between the corrected z -transformation of the upper and lower 95% CI of the ICC coefficients.

All statistical analyses were performed using the IBM SPSS Statistics (Ver. 23 for Windows, IBM Co., Armonk, NY, USA) package. Two-tailed P values less than 0.05 were considered statistically significant.

Results

The LS values assessed by operators with different levels of expertise are shown in Table 1; the values were comparable among the three operators (overall $P=0.980$).

Table 2 shows the LS values assessed by the three operators in various subgroups of subjects. Only the expert operator managed to capture lower LS values in females as compared with the male population ($P=0.028$), while no significant differences between sexes were found by the other two operators ($P=0.225$ for the intermediate and $P=0.140$ for the novice). Higher BMI had no influence on LS measurement, although the difference was near to the significance level for the expert operator ($P=0.058$). No significant differences were found between the first and the last subgroups of 25 subjects for each operator.

An excellent intra-observer agreement was found by the three operators in the overall population ($ICC=0.918$; $P<0.001$) (Table 3). A mild, but progressive reduction of the intra-operator ICC was found with the decline of the operator skill (ICC : 0.941, 0.917, and 0.888 for expert, intermediate and novice operators, respectively) and a significant difference was found between the expert and the novice operators ($P=0.035$) while the intra-observer agreement of the intermediate operator was not significantly different from those of the expert ($P=0.262$) and the novice ($P=0.323$) operators.

As far the role of different factors in influencing the intra-observer agreement is concerned, a significantly better agreement was found in women as compared with men (ICC : 0.930 vs. 0.877; $P=0.029$) (Table 4). This behavior

Table 2 Liver stiffness measurements (mean $\pm$ SD; kPa) of 50 healthy volunteers stratified according to gender, body mass index (BMI) and learning sample

	Expert	Intermediate	Novice	P value*
Gender				
Males	3.98 ± 0.85	3.76 ± 0.70	3.78 ± 0.42	0.179
Females	3.53 ± 0.94	3.55 ± 0.76	3.69 ± 0.92	0.296
P value**	0.028	0.225	0.140	
Body mass index				
BMI < 25 kg/m ²	3.82 ± 0.98	3.63 ± 0.79	3.67 ± 0.79	0.491
BMI > 25 kg/m ²	3.27 ± 0.60	3.61 ± 0.61	3.87 ± 0.74^a	0.046
P value**	0.058	0.785	0.229	
Learning sample				
First 25 subjects	3.69 ± 0.82	3.62 ± 0.68	3.64 ± 0.45	0.961
Last 25 subjects	3.69 ± 1.04	3.63 ± 0.82	3.80 ± 1.00	0.887
P value**	0.432	0.778	0.992	

*Friedman's two-way ANOVA test among the three operators

**Kruskal–Wallis test

^a $P<0.05$ vs. Expert; Friedman's two-way ANOVA

was particularly due to the novice operator which showed a fair grade of agreement in males ($ICC=0.673$); this value was significantly lower than the excellent agreement showed by the same operator in females ($ICC=0.916$; $P=0.005$) as well as it was significantly lower than the agreements showed in males by the other two operators (expert: $ICC=0.915$, $P=0.013$; intermediate: $ICC=0.900$, $P=0.027$). All operators showed an intra-observer agreement lower in overweight than in normal weight subjects; these differences reached the significant level in the overall population ($P<0.001$) and in the expert ($P=0.012$) and the novice ($P=0.022$) operators but not in the intermediate one ($P=0.276$). A significant increase of the agreement was also found with the increasing experience of the operators (ICC : 0.936 vs. 0.881 in the last vs. the former set of subjects, respectively; $P=0.013$). This effect of the experience on

Table 1 Liver stiffness measurements (kPa) of 50 healthy volunteers according to the operators' expertise

Operator	Range	Mean $\pm$ SD	Median (IQR)
Expert	2.46–7.22	3.69 ± 0.93	3.48 (3.12–4.07)
Intermediate	2.60–5.68 ^a	3.63 ± 0.74	3.49 (3.00–4.10)
Novice	2.71–6.56 ^{b,c}	3.72 ± 0.77	3.58 (3.21–4.08)

Friedman's two-way ANOVA: $P=0.980$ among the three operators

^a $P=1.000$ Intermediate vs. Expert

^b $P=0.777$ Intermediate vs. Novice

^c $P=1.000$ Novice vs. Intermediate

Table 3 Intra-observer agreement between different types of operators

	ICC (95% CI)	P value
Overall agreement of the three operators	0.918 (0.897–0.936)	< 0.001
Expert operator	0.941 (0.914–0.963)	< 0.001
Intermediate operator	0.917 (0.877–0.947)	< 0.001
Novice operator	0.888 (0.836–0.929) ^a	< 0.001

^a $P<0.05$ vs. Expert operator

Table 4 Intra-operator variability in the overall sample studied and stratified according to different factors

	All three operators		Expert operator		Intermediate operator		Novice operator	
	ICC (95% CI)	<i>P</i> value	ICC (95% CI)	<i>P</i> value	ICC (95% CI)	<i>P</i> value	ICC (95% CI)	<i>P</i> value
Gender								
Males	0.877 (0.821–0.920)	< 0.001	0.915 (0.841–0.963)	< 0.001	0.900 (0.814–0.956)	< 0.001	0.673 (0.386–0.859) ^{ab}	< 0.001
Females	0.930 (0.907–0.949)	< 0.001	0.950 (0.920–0.972)	< 0.001	0.925 (0.879–0.958)	< 0.001	0.916 (0.865–0.953)	< 0.001
<i>P</i> value	0.029		0.251		0.533		0.005	
Body mass index								
BMI < 25 kg/m ²	0.938 (0.920–0.954)	< 0.001	0.953 (0.927–0.972)	< 0.001	0.927 (0.887–0.957)	< 0.001	0.929 (0.890–0.958)	< 0.001
BMI > 25 kg/m ²	0.818 (0.715–0.895)	< 0.001	0.815 (0.606–0.937)	< 0.001	0.869 (0.722–0.955)	< 0.001	0.744 (0.453–0.914)	< 0.001
<i>P</i> value	< 0.001		0.012		0.276		0.022	
Learning								
First 25 subjects	0.881 (0.837–0.918)	< 0.001	0.927 (0.875–0.963)	< 0.001	0.891 (0.814–0.945)	< 0.001	0.733 (0.544–0.865) ^{ab}	< 0.001
Last 25 subjects	0.936 (0.912–0.956)	< 0.001	0.952 (0.919–0.976)	< 0.001	0.938 (0.894–0.968)	< 0.001	0.921 (0.866–0.960)	< 0.001
<i>P</i> value	0.013		0.340		0.196		0.006	

^a*P* < 0.05 vs. Expert operator^b*P* < 0.05 vs. Intermediate operator**Table 5** Inter-observer agreement between different types of operators

	ICC (95% CI)	<i>P</i> value
Overall agreement among the three operators	0.882 (0.811–0.929)	< 0.001
Expert vs. Intermediate	0.858 (0.751–0.920)	< 0.001
Expert vs. Novice	0.791 (0.631–0.882)	< 0.001
Intermediate vs. Novice	0.849 (0.736–0.914)	< 0.001

the intra-observer agreement was significant in the novice operator only (*ICC*: 0.921 vs. 0.733, *P* = 0.006).

An excellent value of inter-observer agreement for PQE estimation of LS measurement was found among the three raters (overall *ICC* = 0.882; *P* < 0.001) (Table 5). The lowest agreement was seen between the expert and the novice operators (*ICC* = 0.791) although each operator showed no significant differences between the agreement showed with

the other two operators (expert: *P* = 0.350; intermediate: *P* = 0.881; novice: *P* = 0.429). The results of the analysis of the inter-observer reproducibility between pairs of operators are presented in Table 6. The better agreement was found between the intermediate and the novice operators (repeatability coefficient: 1.07 kPa) while the greater variability was found between the expert and the novice ones (repeatability coefficient: 1.40 kPa), thus the variability among operators ranged between ± 1 and ± 1.5 kPa. No significant fixed biases were present while a significant positive proportional bias was found between the expert and the intermediate operators (*P* = 0.019) only. The Bland–Altman plots are showed in Fig. 1. Two outlier values were identified for each of the three pairs of evaluations, but while both of them were within the range of variability of the upper limit of variability as far as the agreement between the expert and the intermediate operators was concerned, one value of the novice operators resulted particularly out of the lower limit of agreement

Table 6 Inter-observer reproducibility between pairs of operators: differences between the measurements made by the couples of operators (kPa)

	Expert vs. Intermediate	Expert vs. Novice	Intermediate vs. Novice
Mean $\pm$ SD (95% CI)	0.064 $\pm$ 0.593 (– 0.100 to 0.229)	– 0.026 $\pm$ 0.713 (– 0.224 to 0.172)	– 0.091 $\pm$ 0.547 (– 0.242 to 0.061)
Range	– 1.071 to 1.539	– 2.735 to 1.652	– 1.881 to 0.958
Repeatability coefficient	1.162	1.398	1.071
Fixed bias	<i>z</i> = 0.768; <i>P</i> = 0.442	<i>z</i> = – 0.259; <i>P</i> = 0.796	<i>z</i> = – 1.171; <i>P</i> = 0.242
Proportional bias	<i>r</i> = 0.331; <i>P</i> = 0.019	<i>r</i> = 0.236; <i>P</i> = 0.098	<i>r</i> = – 0.060; <i>P</i> = 0.678
Lower limit of agreement (95% CI)	– 1.098 (– 1.401 to – 0.914)	– 1.424 (– 1.788 to – 1.203)	– 1.162 (– 1.441 to – 0.993)
Upper limit of agreement (95% CI)	1.227 (1.043 to 1.530)	1.372 (1.151 to 1.736)	0.981 (0.812 to 1.260)

Fig. 1 Bland–Altman plots representing the level of agreement between pairs of operators: **a** expert vs. intermediate, **b** expert vs. novice, **c** intermediate vs. novice. Dark lines show the mean value (solid) of the differences together with its 95% limits of agreement (dashed) while the light dotted lines are the 95% confidence intervals of the lower and upper limits of agreement. The wider is the interval between the limits of agreement, the bigger is the difference between the operators

when compared with both the expert and the intermediate operators.

The role of different factors in influencing the inter-observer agreement is shown in Table 7. No significant differences were found between males and females although the novice operator showed fair agreement values that were only near the significant limit ($ICC=0.576$, $P=0.040$ vs. the expert operator; $ICC=0.656$, $P=0.020$ vs. the intermediate

operator). On the other hand, the inter-observer agreement was significantly lower in overweight/obese subjects than in normal weight subjects (ICC : 0.603 vs. 0.923; $P=0.011$). In particular, in overweight/obese subjects, the overall agreement among the three operators ($ICC=0.603$, $P=0.016$), as well as the agreement between the expert and the intermediate operator ($ICC=0.692$, $P=0.016$), resulted only near the significant level while the novice operator showed no significant agreement with both the expert ($ICC=0.263$, $P=0.255$) and the intermediate ($ICC=0.576$, $P=0.079$) operators. Thus, the agreement of the novice operator was greatly influenced by the BMI: the agreement with expert operators was significantly lower in overweight/obese subjects than in normal weight subjects (ICC : 0.263 vs. 0.883, $P=0.021$) while the comparison between overweight/obese ($ICC=0.576$) and normal weight ($ICC=0.902$) subjects of the agreement showed by the novice operator with the expert operator was only near the significant level ($P=0.092$). Finally, the learning has no effect on the inter-observer agreement; in fact no significant differences were found between the agreement observed in the evaluation of the first 25 subjects and the agreement observed in the last 25 subjects (Table 7).

The intra- and inter-operator agreements obtained using the two subsets of five measurements (i.e., first five vs. last five) are shown in Table 8. The ICC values in the two subsets of five measurements are lower than those obtained using the entire set of ten measurements. Both sets of five measurements present an overall intra-operator agreement significantly lower than that of the ten measurements. No significant differences were found as far the inter-operator agreement was concerned. The reduction of the intra-operator agreement by considering five instead of ten measurements was significant in the subjects with normal weight only, while no significant reduction was found in the overweight subjects.

Discussion

The main objective of this study was to assess the good performance and reproducibility of PQE in normal subjects by operators with different expertise. These findings are important and clinically relevant, because nowadays the use of elastographic/elastometric techniques is widening and almost always implied, in non-invasive assessment of CLD. Unlike TE, which is a specially developed device for measuring LS, PQE is embedded into an US machine, and so presumably previous knowledge on ultrasound manipulation is expected. Our study contradicts this rationale, demonstrating that even inexperienced US operators can obtain reliable results, comparable with expert and intermediate operators. In our study, an excellent inter-operator agreement

Table 7 Inter-observer agreement among the three operators according to different factors

	Among the three operators		Expert vs. Intermediate		Expert vs. Novice		Intermediate vs. Novice	
	ICC (95% CI)	<i>P</i> value	ICC (95% CI)	<i>P</i> value	ICC (95% CI)	<i>P</i> value	ICC (95% CI)	<i>P</i> value
Gender								
Males	0.783 (0.531–0.912)	< 0.001	0.810 (0.508–0.928)	< 0.001	0.576 (– 0.089–0.893)	0.040	0.656 (0.049–0.873)	0.020
Females	0.908 (0.836–0.952)	< 0.001	0.876 (0.745–0.940)	< 0.001	0.843 (0.681–0.923)	< 0.001	0.891 (0.779–0.947)	< 0.001
<i>P</i> value	0.118		0.520		0.147		0.101	
Body mass index								
BMI < 25 kg/m ²	0.923 (0.867–0.957)	< 0.001	0.882 (0.768–0.940)	< 0.001	0.883 (0.774–0.939)	< 0.001	0.902 (0.812–0.949)	< 0.001
BMI > 25 kg/m ²	0.603 (0.069–0.869)	0.016	0.692 (0.037–0.909)	0.016	0.263 (– 0.638–0.744)	0.255	0.576 (– 0.348–0.875)	0.079
<i>P</i> value	0.011		0.233		0.021		0.092	
Learning								
First 25 subjects	0.864 (0.736–0.936)	< 0.001	0.880 (0.729–0.947)	< 0.001	0.727 (0.376–0.880)	0.001	0.781 (0.498–0.904)	< 0.001
Last 25 subjects	0.893 (0.792–0.949)	< 0.001	0.848 (0.655–0.933)	< 0.001	0.821 (0.594–0.921)	< 0.001	0.877 (0.725–0.946)	< 0.001
<i>P</i> value	0.509		0.692		0.487		0.338	

among the three operators was found ($ICC=0.882$), as well as between each two categories at a time (Table 5). These data are concurrent with other reports that evaluated the agreement between expert and novice examinations using different elastographic techniques in patients with CLD or healthy subjects: ICC of 0.82 for the intercostal assessment with point SWE in patients with chronic hepatitis [25], or ICC of 0.88 in healthy subjects evaluated by 2D-SWE [26]. This finding are also in line with previous data about the reliability of Fibroscan in hands of a non-experienced operator [27] and opens the possibility of LS assessment—with various techniques, by non-specialists (trained nurses or doctors without US background).

With the advent of numerous shear wave-based elastographic techniques, the determination of normal LS range is an important prerequisite for clinical practice. Several studies in healthy population have investigated normal LS by either TE or ARFI [28–32]. The mean stiffness value of normal liver in different studies was 5.49 kPa (range 1.5–12.7 kPa) [30], 4.6 kPa (range 3.3–5.6 kPa) [31], and 4.4 kPa (range 2.1–17.5 kPa) [32]. Our study revealed a slightly lower median stiffness value for patients without liver disease: 3.5 kPa (range 2.47–7.16 kPa), with the tendency of differentiation according to gender: in men, normal liver seem to be stiffer than in women, even if this finding reached the significance in the expert operator only. This is concurrent with previous findings from a head-to-head comparison of elastographic techniques in chronic viral hepatitis patients [33] and with feasibility data of PQE in Asian healthy subjects [34].

Obesity ($BMI > 30 \text{ kg/m}^2$) was identified as the most important factor associated with LS measurement failure at TE, and in the same large French cohort, the proportion

of unreliable results reached almost 25% in patients with $BMI > 30 \text{ kg/m}^2$ [35]. Taking it into account, we wanted to evaluate the influence of BMI on the LS values assessed by PQE. In all of our moderately overweight subjects LS measurement by PQE was feasible, with no particular difficulties for less experienced examiners, demonstrating the importance of having a visual control. In our population, we found no differences in LS values among subjects with $BMI > 25 \text{ kg/m}^2$ and those with $BMI < 25 \text{ kg/m}^2$ (Table 2), but the level of agreement was significantly lower in overweight subjects than in normal weight ones (Table 7). If we consider the small number of subjects together with the selection bias (healthy volunteers), and the slight elevated BMI, severe obesity cannot yet be excluded as a drawback for PQE quantification, further investigation are needed with “real obese patients”.

We also wanted to assess the learning curve of PQE, since one of our operators had no experience in the use of ultrasound. Considering all the three operators we did not find any difference among the liver stiffness values acquired from the first 25 subjects and the last 25 ones, neither among the ICC. Considering the novice operator we found a better agreement in the last 25 subjects, suggesting a kind of training effect in non-skilled operators.

We found an excellent intra-observer agreement for each operator with the tendency to increase with the level of expertise (Table 3). The finding is also in trend with previous reports [13]. Both BMI and gender significantly influence the performance of the reproducibility in the case of the novice operator, demonstrating the role of US experience in overcoming some challenges related to PQE examinations.

Finally, a reduction in the number of acquisitions from 10 to 5 led to a significant reduction of intra-operator agreement

Table 8 Comparison of intra- and inter-operator variability calculated by taking into account two sets of five measurements (the first five and the last five measurements within the overall set of ten measure-

ments) vs. the intra- and inter-operator variability calculated in the entire set of ten measurements

	First five measurements			Last five measurements		
	ICC (95% CI)	Δ^a	<i>P</i> value ^a	ICC (95% CI)	Δ^a	<i>P</i> value ^a
Overall population						
All three operators ^b	0.866 (0.829–0.897)	– 0.052	0.006	0.853 (0.812–0.887)	– 0.065	0.001
Expert ^b	0.926 (0.888–0.954)	– 0.015	0.474	0.874 (0.808–0.922)	– 0.067	0.017
Intermediate ^b	0.834 (0.749–0.897)	– 0.083	0.028	0.863 (0.793–0.915)	– 0.054	0.112
Novice ^b	0.761 (0.639–0.852)	– 0.127	0.016	0.833 (0.747–0.897)	– 0.055	0.206
Normal weight subjects						
All three operators ^c	0.903 (0.872–0.929)	– 0.035	0.031	0.895 (0.862–0.923)	– 0.043	0.011
Expert ^c	0.949 (0.918–0.971)	– 0.004	0.829	0.899 (0.837–0.942)	– 0.054	0.035
Intermediate ^c	0.859 (0.772–0.919)	– 0.068	0.071	0.874 (0.798–0.928)	– 0.053	0.134
Novice ^c	0.837 (0.738–0.906)	– 0.092	0.022	0.912 (0.859–0.950)	– 0.017	0.559
Overweight subjects ^c						
All three operators ^c	0.647 (0.425–0.801)	– 0.171	0.080	0.676 (0.474–0.817)	– 0.142	0.126
Expert ^c	0.662 (0.224–0.889)	– 0.153	0.394	0.649 (0.185–0.885)	– 0.166	0.368
Intermediate ^c	0.714 (0.350–0.906)	– 0.155	0.263	0.819 (0.585–0.941)	– 0.050	0.648
Novice ^c	0.360 (–0.531–0.794)	– 0.384	0.254	0.591 (0.048–0.866)	– 0.153	0.516
Overall population						
Among the three operators ^d	0.814 (0.702–0.888)	– 0.068	0.199	0.860 (0.776–0.916)	– 0.022	0.629
Expert vs. Intermediate ^d	0.760 (0.576–0.864)	– 0.098	0.205	0.837 (0.713–0.907)	– 0.021	0.737
Expert vs. Novice ^d	0.693 (0.457–0.826)	– 0.098	0.356	0.744 (0.548–0.855)	– 0.047	0.625
Intermediate vs. Novice ^d	0.794 (0.637–0.883)	– 0.055	0.448	0.832 (0.705–0.904)	– 0.017	0.793
Normal weight subjects						
Among the three operators ^e	0.853 (0.749–0.918)	– 0.070	0.112	0.912 (0.850–0.951)	– 0.011	0.742
Expert vs. Intermediate ^e	0.771 (0.563–0.880)	– 0.111	0.170	0.882 (0.770–0.939)	0.000	1.000
Expert vs. Novice ^e	0.781 (0.582–0.886)	– 0.102	0.188	0.850 (0.713–0.922)	– 0.033	0.600
Intermediate vs. Novice ^e	0.855 (0.720–0.925)	– 0.047	0.411	0.890 (0.789–0.943)	– 0.012	0.807
Overweight subjects ^e						
Among the three operators ^e	0.543 (– 0.022–0.845)	– 0.060	0.844	0.516 (– 0.186–0.844)	– 0.087	0.793
Expert vs. Intermediate ^e	0.718 (0.043–0.919)	0.026	0.924	0.512 (– 0.529–0.855)	– 0.180	0.638
Expert vs. Novice ^e	0.181 (– 0.582–0.694)	– 0.082	0.883	0.165 (– 1.288–0.737)	– 0.098	1.000
Intermediate vs. Novice ^e	0.378 (– 0.902–0.814)	– 0.198	0.746	0.566 (– 0.480–0.874)	– 0.010	0.982

The data are reported in the overall population and by considering the three operators pooled together, as well as they are stratified according to BMI and the single operators

^aDifferences and *P* values vs. ICC evaluated using all of ten measurements

^bThe ICC values of all of ten measurements are shown in Table 6

^cThe ICC values of all of ten measurements are shown in Table 7

^dThe ICC values of all of ten measurements are shown in Table 4

^eThe ICC values of all of ten measurements are shown in Table 5

in normal weight but not in overweight subject, and this is due the fact that, in these patients, a wider variability and a lower agreement was observed also on ten measurements for all operators. On the other hand, no significant differences in intra-operator agreement were found. Even if some previous studies suggested that a valid assessment of LS could be obtained with less than ten measurements, according to our results, we suggest acquiring no less than ten valid

measurements. This is considered also that our results were obtained in a population of healthy volunteers in which the LS values are not affected by pathological factors that clearly influence measurement variability.

In summary, this study demonstrates the wide applicability of PQE technique, which has excellent intra- and inter-operator agreement, preserving high reproducibility and reliability even in the hands of less or inexperienced operators.

Despite this, we believe that a basic ultrasound knowledge is desirable to minimize intra-operator variability, especially in obese or technically difficult patients. The level of agreement with expert operator tends to decrease in case of overweight subjects, and also in men. Based on excellent intra-observer agreement figures, one can hypothesize that five measurements would be enough to have an adequate estimation of LS, apart from the novice operator examining overweight patients; in this case, we suggest collecting ten live stiffness values.

Author contributions All the authors have made a significant contribution to this manuscript, have seen and approved the final manuscript, and have agreed to its submission to the Journal of Ultrasound.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval The study protocol conforms to the ethical guidelines of the World Medical Association Declaration of Helsinki; all patients gave their written consent and the study has been notified to the institutional review board of the St Orsola-Malpighi University Hospital.

Informed consent Informed consent was obtained for all individual participants included in the study.

References

- Bravo AA, Sheth SG, Chopra S (2001) Liver biopsy. *N Engl J Med* 344(7):495–500
- Piccinino F, Sagnelli E, Pasquale G, Giusti G (1986) Complications following percutaneous liver biopsy. A multicentre retrospective study on 68,276 biopsies. *J Hepatol* 2(2):165–173
- Bedossa P, Dargere D, Paradis V (2003) Sampling variability of liver fibrosis in chronic hepatitis C. *Hepatology* 38(6):1449–1457
- Bamber J, Cosgrove D, Dietrich CF, Fromageau J, Bojunga J, Calliada F, Cantisani V, Correas JM, D'Onofrio M, Drakonaki EE, Fink M, Friedrich-Rust M, Gilja OH, Havre RF, Jenssen C, Klauser AS, Ohlinger R, Saftoiu A, Schaefer F, Sporea I, Piscaglia F (2013) EFSUMB guidelines and recommendations on the clinical use of ultrasound elastography. Part 1: Basic principles and technology. *Ultraschall Med* 34:169–184 **EFSUMB Guidelines 1**
- Cosgrove D, Piscaglia F, Bamber J, Bojunga J, Correas JM, Gilja OH, Klauser AS, Sporea I, Calliada F, Cantisani V, D'Onofrio M, Drakonaki EE, Fink M, Friedrich-Rust M, Fromageau J, Havre RF, Jenssen C, Ohlinger R, Saftoiu A, Schaefer F (2013) Dietrich CF; EFSUMB. EFSUMB guidelines and recommendations on the clinical use of ultrasound elastography. Part 2: clinical applications. *Ultraschall Med* 34(3):238–253. <https://doi.org/10.1055/s-0033-1335375> **Epub 2013 Apr 19 PubMed PMID: 23605169**
- Shiina T, Nightingale KR, Palmeri ML, Hall TJ, Bamber JC, Barr RG, Castera L, Choi BI, Chou YH, Cosgrove D, Dietrich CF, Ding H, Amy D, Farrokhi A, Ferraioli G, Filice C, Friedrich-Rust M, Nakashima K, Schafer F, Sporea I, Suzuki S, Wilson S, Kudo M (2015) WFUMB guidelines and recommendations for clinical use of ultrasound elastography. *Ultrasound Med Biol* 41(5):1126–1147. <https://doi.org/10.1016/j.ultrasmedbio.2015.03.009> **Epub 2015 Mar 21. Review**
- Ferraioli G, Filice C, Castera L, Choi BI, Sporea I, Wilson SR, Cosgrove D, Dietrich CF, Amy D, Bamber JC, Barr R, Chou YH, Ding H, Farrokhi A, Friedrich-Rust M, Hall TJ, Nakashima K, Nightingale KR, Palmeri ML, Schafer F, Shiina T, Suzuki S, Kudo M (2015) WFUMB guidelines and recommendations for clinical use of ultrasound elastography: part 3: liver. *Ultrasound Med Biol* 41(5):1161–1179. <https://doi.org/10.1016/j.ultrasmedbio.2015.03.007> **Epub 2015 Mar 20 PubMed PMID: 25800942**
- Dietrich CF, Bamber J, Berzigotti A, Bota S, Cantisani V, Castera L, Cosgrove D, Ferraioli G, Friedrich-Rust M, Gilja OH, Goertz RS, Karlas T, de Knecht R, de Ledinghen V, Piscaglia F, Procopet B, Saftoiu A, Sidhu PS, Sporea I, Thiele M (2017) EFSUMB guidelines and recommendations on the clinical use of liver ultrasound elastography, update 2017 (long version). *Ultraschall Med* 38(4):e16–e47. <https://doi.org/10.1055/s-0043-103952> **Epub 2017 Apr 13 PubMed PMID: 28407655**
- Ziol M, Handra-Luca A, Kettaneh A, Christidis C, Mal F, Kazemi F, de Ledinghen V, Marcellin P, Dhumeaux D, Trinchet JC, Beaugrand M (2005) Noninvasive assessment of liver fibrosis by measurement of stiffness in patients with chronic hepatitis C. *Hepatology* 41(1):48e54
- Castéra L, Vergniol J, Foucher J, Le Bail B, Chanteloup E, Haaser M, Darriet M, Couzigou P, De Ledinghen V (2005) Prospective comparison of transient elastography, fibrotest, APRI, and liver biopsy for the assessment of fibrosis in chronic hepatitis C. *Gastroenterology* 128(2):343e50
- Arena U, Vizzutti F, Abrandes JG, Corti G, Stasi C, Moscarella S, Milani S, Lorefice E, Petrarca A, Romanelli RG, Laffi G, Bosch J, Marra F, Pinzani M (2008) Reliability of transient elastography for the diagnosis of advanced fibrosis in chronic hepatitis C. *Gut* 57(9):1288e93
- Wong VW, Vergniol J, Wong GL, Foucher J, Chan HL, Le Bail B, Choi PC, Kow M, Chan AW, Merrouche W, Sung JJ, de Ledinghen V (2010) Diagnosis of fibrosis and cirrhosis using liver stiffness measurement in nonalcoholic fatty liver disease. *Hepatology* 51(2):454e62
- Boursier J, Konate A, Guilluy M, Gorea G, Sawadogo A, Quemenner E, Oberti F, Reaud S, Hubert-Fouchard I, Dib N, Calès P (2008) Learning curve and interobserver reproducibility evaluation of liver stiffness measurement by transient elastography. *Eur J Gastroenterol Hepatol* 20(7):693–701. <https://doi.org/10.1097/MEG.0b013e3282f51992>
- Tsochatzis EA, Gurusamy KS, Ntaoula S, Cholongitas E, Davidson BR, Burroughs AK (2011) Elastography for the diagnosis of severity of fibrosis in chronic liver disease: a meta-analysis of diagnostic accuracy. *J Hepatol* 54:650–659. <https://doi.org/10.1016/j.jhep.2010.07.033>
- Coco B, Oliveri F, Maina AM et al (2007) Transient elastography: a new surrogate marker of liver fibrosis influenced by major changes of transaminases. *J Viral Hepat* 14(5):360–369
- Ferraioli G, Tinelli C, Lissandrin R, Zicchetti M, Dal Bello B, Filice G, Filice C (2014) Point shear wave elastography method for assessing liver stiffness. *World J Gastroenterol* 20(16):4787–4796. <https://doi.org/10.3748/wjg.v20.i16.4787>

17. Goertz RS, Zopf Y, Jugl V, Heide R, Janson C, Strobel D, Bernatik T, Haendl T (2010) Measurement of liver elasticity with acoustic radiation force impulse (ARFI) technology: an alternative noninvasive method for staging liver fibrosis in viral hepatitis. *Ultraschall Med* 31(2):151–155
18. Nightingale K, McAleavey S, Trahey G (2003) Shear-wave generation using acoustic radiation force: in vivo and ex vivo results. *Ultrasound Med Biol* 29(12):1715–1723
19. Sporea I, Grădinaru-Tașcău O, Bota S, Popescu A, Șirli R, Jurchiș A, Popescu M, Dănilă M (2013) How many measurements are needed for liver stiffness assessment by 2D-Shear Wave Elastography (2D-SWE) and which value should be used: the mean or median? *Med Ultrason* 15(4):268–272
20. Fleiss JL (1986) The design and analysis of clinical experiments. Wiley, New York
21. Bland JM, Altman DG (2003) Applying the right statistics: analyses of measurement studies. *Ultrasound Obstet Gynecol* 22:85–93
22. Carkeet A (2015) Exact parametric confidence intervals for Bland–Altman limits of agreement. *Optom Vis Sci* 92(3):e71–e80
23. British Standards Institution (1975) Precision of test methods 1: guide for the determination and reproducibility for a standard test method (BS 5497, part 1). BSI, London
24. Fisher RA (1934) Statistical methods for research workers. 5th edn. Oliver and Boyd, Edinburgh, pp 203–209
25. Ferraioli G, Tinelli C, Lissandrin R, Zicchetti M, Dal Bello B, Filice C (2013) Performance of ElastPQ shear wave elastography technique for assessing fibrosis in chronic viral hepatitis (abstract). *J Hepatol* 58(1):S7
26. Ferraioli G, Tinelli C, Zicchetti M, Aboe E, Poma G, Di Gregorio M, Filice C (2012) Reproducibility of real-time shear wave elastography in the evaluation of liver elasticity. *Eur J Radiol* 81(11):3102
27. McCorry RB, Palaniyappa N, Chivinge A, Kaye P, James MW, Aithal GP (2012) Development and evaluation of a nurse-led transient elastography service for the staging of hepatic fibrosis in patients with suspected chronic liver disease. *Q J Med* 105:749–754
28. D’Onofrio M, Gallotti A, Mucelli RP (2010) Tissue quantification with acoustic radiation force impulse imaging: measurement repeatability and normal values in the healthy liver. *Am J Roentgenol* 195:132–136
29. Guzmán-Aroca F, Reus M, Berná-Serna JD et al (2011) Reproducibility of shear wave velocity measurements by acoustic radiation force impulse imaging of the liver: a study in healthy volunteers. *J Ultrasound Med* 30:975–979
30. Roulot D, Czernichow S, Le Clesiau H, Costes JL, Vergnaud AC, Beaugrand M (2008) Liver stiffness values in apparently healthy subjects: influence of gender and metabolic syndrome. *J Hepatol* 48(4):606–613
31. Kim SU, Choi GH, Han WK, Kim BK, Park JY, Kim DY, Choi JS, Yang SC, Choi EH, Ahn SH, Han KH, Chon CY (2010) stiffness values using FibroScan?: a prospective study in healthy living liver and kidney donors in South Korea. *Liver Int* 30(2):268–274
32. Colombo S, Belloli L, Zaccanelli M, Badia E, Jamoletti C, Buonocore M, Del Poggio P (2011) Normal liver stiffness and its determinants in healthy blood donors. *Dig Liver Dis* 43(3):231–236
33. Ling W, Lu Q, Quan J, Ma L, Luo Y (2013) Assessment of impact factors on shear wave based liver stiffness measurement. *Eur J Radiol* 82(2):335–341
34. Kim BK, Kim SU, Choi GH, Han WK, Park MS, Kim EH, Park JY, Kim DY, Choi JS, Yang SC, Choi EH, Song K, Ahn SH, Han KH, Chon CY (2012) “Normal” liver stiffness values differ between men and women: a prospective study for healthy liver and kidney donors in a native Korean population. *J Gastroenterol Hepatol* 27(4):781–788
35. Castéra L, Foucher J, Bernard PH, Carvalho F, Allaix D, Merrouche W, Couzigou P, de Ledinghen V (2010) Pitfalls of Liver stiffness measurement: a 5 year prospective study of 13,369 examinations. *Hepatology* 51:828–835

Affiliations

Cristina Felicani¹ · Chiara De Molo¹ · Horia Stefanescu^{2,3} · Fabio Conti³ · Elena Mazzotta¹ · Veronica Gabusi¹ · Elena Nardi³ · Antonio Maria Morselli-Labate³ · Pietro Andreone³ · Carla Serra¹

✉ Cristina Felicani
cristinafelicani@gmail.com

Chiara De Molo
chiara.demolo@gmail.com

Horia Stefanescu
hstfnc@gmail.com

Fabio Conti
fabio.conti2@studio.unibo.it

Elena Mazzotta
elena.mazzotta@aosp.bo.it

Veronica Gabusi
veronicagabusi2@gmail.com

Elena Nardi
elena.nardi2@unibo.it

Antonio Maria Morselli-Labate
antonio.morselli@unibo.it

Pietro Andreone
pietro.andreone@unibo.it

Carla Serra
carla.serra@aosp.bo.it

¹ Interventional, Diagnostic and Therapeutic Ultrasound Unit, St Orsola-Malpighi Hospital, University of Bologna, Bologna, Italy

² Department of Hepatology, Regional Institute of Gastroenterology and Hepatology, Cluj-Napoca, Romania

³ Department of Medical and Surgical Sciences, St Orsola-Malpighi Hospital, University of Bologna, Bologna, Italy