Introduction

Ultrasound elastography is increasingly being used in diagnostics and follow-up of liver diseases in children [1,2], and could in some cases make liver biopsy obsolete in the evaluation of liver fibrosis development. A sensitivity and specificity of 81% and 91% for the detection of liver fibrosis has been reported [2], and the technique is continuously improving. Potentially, it can help to avoid other invasive procedures, such as endoscopy: The Baveno VI criteria for surveillance for esophageal varices in patients with compensated cirrhosis, cite liver stiffness measurement <20 kPa as one of the criteria for avoiding endoscopy [3]. Such implications would be especially welcomed in pediatrics, as both liver biopsy and endoscopy often necessitate the use of general anesthesia. Liver biopsy is still the criterion standard, but is challenged due to invasiveness, need for general anesthesia and the potential of sampling errors and complications [4]. Furthermore, liver biopsy is not part of the primary investigations in children with low pretest probability of serious liver disease.

The number of recommended acquisitions for a valid LSM value is debated, but for the purpose of efficiency in clinical practice, particularly when it comes to examination of children, knowledge about the lower threshold for number of repeated elastography acquisitions for the different methods is important. The current EFSUMB guidelines recommend 10 acquisitions for pSWE and TE and only 3 for 2D-SWE [5], and the basic principles and technology of the different systems are explained as part of the same guidelines [6]. However, these and other recommendations reflect expert opinion based on a mix between producer recommendations and currently published investigations. Despite the current consensus regarding the number of acquisitions needed, some reports argue that a sufficient number of acquisitions might be 3 for TE [7] and 5 or 6 for pSWE [8,9]. The sufficiency of only 3 acquisitions for 2D-SWE, has also been challenged [10]. That different systems yield different values, however, is well documented [11-13].

We aimed to evaluate the number of acquisitions needed to produce a reliable LSM in healthy children, investigating the difference between the median of 3-8 acquisitions and the median of 10 acquisitions. We wanted to determine the number of acquisitions needed for LSM to achieve an ICC with absolute agreement >0.95 when compared to LSM based on 10 acquisitions, without individual LSM deviating above 20% from the median of ten acquisitions. Furthermore, we wanted as few as possible (<10%) of LSM deviating 10% or more.

Material and methods

Study design and subjects

The study was designed as a single-center prospective study. Participants were recruited through hospital employees, social media and invitations submitted to local schools. All children and their legal guardians were given oral and written information, and informed written consent was obtained. Exclusion criteria were a history of liver disease, and other diseases, or the use of medication, with known potential of affecting the liver. A total of 246 healthy children (111 boys; 45.1%) between 4 and 17 years were recruited. Subjects who had not fasted (n=1) or who showed B-mode signs of steatosis or splenomegaly (n=2) were excluded, leaving 243 for analysis. Participants were divided into four age predefined age categories: 4-7; 8-11; 12-14 and 15-17 years.

The medical history was obtained, and all subjects underwent a clinical evaluation by an experienced pediatrician, including measuring height and weight, prior to ultrasound examination.

B-mode ultrasound examination

B-mode ultrasound was performed according to a standardized protocol, examining liver, spleen and kidneys, using Samsung RS80A with Prestige, with a convex 1-7 MHz probe.

Liver stiffness measurements (LSM)

Liver stiffness measurements were performed with the subject in a horizontal position, with right arm maximally abducted, after ≥3 hours of fasting. The transducer was placed in a right intercostal space, perpendicular to the surface, with the region of interest (ROI) 2-5 cm below the liver capsule, avoiding vessels and visible bile ducts. Acquisitions were performed during mid-expiratory breath-hold, or if not possible, during calm expiration. In all subjects, both point shear wave elastography (pSWE, Samsung RS80A with Prestige), using a CA1-7A probe, and two-dimensional shear wave elastography (2D-SWE, GE Logiq E9), using a C1-6 probe, was performed, in a head-to-head fashion. In a subset among participants aged 8-17 years (n=87), transient elastography (TE, Fibroscan, incorporated in a GE S8, GE Healthcare) was performed. For all systems, the same intercostal space was used, and for 2D-SWE, the ROI was placed in a color elastogram without automatic LSM display. All measurements were performed by a single operator (A.B.M.) with more than 2 years' experience in abdominal ultrasound and liver stiffness measurements using pSWE, 2D-SWE and TE, and was certified as a Fibroscan user. The operator was not blinded to the specific LSM results. Ten acquisitions were performed, with valid measurements requiring an interquartile range divided by the median (IQR/M) less than 30% and a success rate ≥60%.

Statistical analyses

We used SPSS version 25 (SPSS Inc, 2016, Armonk, NY) for all analyses. Median values for the first three to eight acquisitions (LSM_3, LSM_4, LSM_5, LSM_6, LSM_7, LSM_8) were made across all three systems. All median values were compared with the median of ten acquisitions, which was considered reference standard. Variables were tested for normality using the Shapiro-Wilk test and a Q-Q plot, and presented as mean (SD) or median (minimum, maximum), as deemed appropriate. Correlations were tested by Pearson correlation coefficient and intraclass correlation coefficients (ICCs), with absolute agreement, between median values were calculated. Limits of agreement were assessed

to evaluate the reliability and the possibility of bias. We used Bland-Altman plots for the evaluation of ICCs.

Ethical aspects

The protocol was in accordance with the Declaration of Helsinki and approved by the Regional Committee on Medical and Health Research Ethics (2017/290).

Results

The characteristics of the 243 participants are found in Table 1. LSM varied across age groups; thus, we performed comparisons separately for every age group as well as for all participants combined. LSM results for all medians are displayed in Table 2 and 3. Valid results were obtained in 238/243 (97.9%) for pSWE, 242/243 (99.6%) for 2D-SWE and 83/87 (95.4%) for TE. Invalid results were due to an IQR/M above 30%.

Across all age groups and systems, there was no significant difference between medians of three, four, five, six, seven and eight, respectively, and median of all ten acquisitions (all p-values ≥0.4). Similarly, there was no significant difference when looking at the entire cohort together.

LSM values for the medians of 3-8 values, and for the individual age groups, are shown in Table 2 and 3, respectively.

LSM by point shear wave elastography, pSWE (Samsung RS80A with Prestige)

Median of three acquisitions (LSM_3) showed an excellent ICC of 0.97 (95% CI: 0.959-0.978); only 2 (0.8%) were above 20% deviant from the LSM_{10}, and 201 (84.4%) were within 10% difference (Table 2). The highest absolute difference between LSM_3 and LSM_{10} was 1.2 kPa, the highest relative difference -20.5%. The number of acquisitions to achieve zero LSM ≥20% from LSM_{10} was 4, with an ICC of 0.98.

LSM by two-dimensional shear wave elastography, 2D-SWE (GE Logiq E9)

LSM_3 showed an excellent ICC of 0.98 (95% CI: 0.975-0.985), with 10 (4.1%) ≥20% deviant from LSM_{10}, and 206 (85.1%) within 10% difference (Table 2). The highest

absolute difference between LSM_3 and LSM_{10} was 1.3 kPa, the highest relative difference 50%. The number of acquisitions to achieve no LSM ≥20% from LSM_{10} was 6, yielding an ICC of 0.996.

LSM by transient elastography, TE (Fibroscan)

LSM_3 showed an excellent ICC of 0.98 (95% CI: 0.961-0.984), with 5 (6.1%) ≥20% deviant from LSM_{10}, and 57 (69.5%) within 10% difference (Table 2). The highest absolute difference between LSM_3 and LSM_{10} was 1.0 kPa, the highest relative difference 27%. The number of acquisitions to achieve no LSM ≥20% from LSM_{10} was 4, with an ICC of 0.98.

Comparing the different elastography systems

We found that all systems showed excellent correlation between LSM_4 and LSM_{10} (Fig. 1a-c). However, the percentage of median values (LSM_{3-8}) deviating more than 10% or 0.5 kPa from LSM_{10} fell rapidly from 3 towards 6 acquisitions, before flattening out (Fig. 2a-b). The number of required acquisitions depended on the chosen quality criterion and varied between systems for some of the criteria (Table 4).

Dividing subjects in those with highly reliable measurements (IQR/M <10%) or moderately reliable measurements (IQR/M 20-30%), the ICC was markedly lower for the latter (Table 5). For the moderately reliable across all systems, ≥8 acquisitions were required to achieve similar ICC values as for LSM_3 in subjects with a low IQR/M.

We could not find any correlation between LSM differences and age, weight, BMI or gender.

Discussion

Liver elastography is established as a highly useful tool in diagnostics and follow-up of chronic liver diseases in children; however, the optimal number of acquisitions is debated. The current international guidelines recommend ten acquisitions for pSWE and TE, whilst only three for 2D-SWE [5], but underlying publications demonstrate conflicting

results. This question is clinically relevant and important, as less time spent on acquisitions, increases workflow and reduces time spent at the institution for the patient. In children, it is imperative not to use more time and effort than necessary, as increased time spent on examination may lead to reduced cooperation from the child, making it difficult to obtain valid measurements.

We found that for all systems, 3 acquisitions yielded an ICC above 0.97 (using absolute agreement) for the corresponding LSM compared to the reference LSM based on 10 acquisitions. However, application of ICC as a quality criterion alone allowed clinically important deviations in LSM (≥1 kPa) from the reference in individual subjects. Thus, 4 acquisitions resulted in no LSM deviating 1 kPa or more from the reference standard. To achieve zero LSM with a difference ≥20% of LSM_{10}, the required amounts of acquisitions were 4, 4 and 6 for pSWE, TE and 2D-SWE, respectively. Hence, we could not confirm previous assumptions that 2D-SWE requires less acquisitions than pSWE or TE.

We prepared Bland-Altman plots comparing medians of 3 vs. 10 (Fig. 3a-c) and 5 vs. 10 acquisitions (Fig. 3d-f), suggesting that pSWE had a lower degree of variation compared to 2D-SWE and TE when using only 3 acquisitions, while this difference disappeared when 5 acquisitions were performed. Although we cannot conclude with certainty from this, our results indicate a need for reevaluation of the recommendation of 10 acquisitions for pSWE and only 3 for 2D-SWE.

Previous studies have reported that 3 or 5 acquisitions with TE show similar performance for the diagnosis of cirrhosis, compared with 10 acquisitions [7], while others have suggested that 6 acquisitions may be optimal in liver patients, both for pSWE [9] and 2D-SWE [10]. Other studies comparing either 3 or 5 acquisitions with 2D-SWE [14,15], found no difference. One study showed no difference between 5 and 10 acquisitions in healthy adults for either pSWE and 2D-SWE [8], while a study in both healthy and subjects with liver disease showed more reliable results with 10 as compared to 5 acquisitions using pSWE [16], needing 8 acquisitions to reach <5% deviation from the median of 10 acquisitions.

It is conceivable that there is an increased need of acquisitions in fibrotic compared to healthy livers: as fibrosis often is unevenly distributed, this could increase the dispersion of values from the individual acquisitions. Therefore, our finding that 4 acquisitions are sufficient for a reliable LSM across all platforms is most relevant in the setting of screening for liver diseases in children, where the pretest probability is rather low. This is supported by our finding of higher ICC in subjects with a low IQR/M.

Our finding that 2D-SWE needed more acquisitions to achieve no LSM ≥20% from the reference standard, as compared to TE and pSWE, could be due to the fact that 2D-SWE has significantly lower absolute LSM values, making identical absolute differences in kPa relatively larger for 2D-SWE. This is reflected by the fact that the number of acquisitions needed to avoid differences ≥1.0 kPa, was identical in all systems. Thus, we believe that a reasonable conclusion is that the number of acquisitions needed for a reliable LSM is similar across these three systems. It should, however, be mentioned that publications on 2D-SWE cited by current guidelines, are based on one specific system: supersonic shear imaging (SSI; Aixplorer, Supersonic Imagine, France) [10,14,15,17-20].

Application of more rigid reliability criteria (i.e. an IQR/M <10%) did not significantly improve the *difference* between median values; however, the ICC improved (Table 5). It has previously been suggested that IQR/M should be divided into very reliable (<10%), reliable (10-30%) and poorly reliable (>30%) [21], but guidelines only separate into valid (≤30%) or invalid measurements (>30%). It should be noted that the IQR/M limit of 30% is solely valid when performing LSM in kilopascals, while a limit close to 15% should be sought when measuring LSM in m/s [22]. A low IQR/M could however support the investigator's confidence in measurements made with few acquisitions, and it is reasonable to postulate that more acquisitions are needed in fibrotic compared to healthy livers. Furthermore, this strengthens the foundation for common clinical practice: to continue acquisitions when one observes that more acquisitions lowers IQR/M.

The study would have been strengthened if children with chronic liver disease had been included, and similar studies in such a cohort are warranted.

Conclusion

Our findings in this head-to-head investigation of pSWE, 2D-SWE and TE in a healthy pediatric cohort suggest that four acquisitions are suitable for a valid LSM across all systems. We believe that this number may be suitable when screening for liver disease in children, contradicting international guidelines recommending only three acquisitions for 2D-SWE compared to ten acquisitions for pSWE and TE, but similar studies in children with chronic liver disease are warranted. Both pretest probability of liver disease as well as IQR/M should be taken into consideration when performing acquisitions.

References.

1. Andersen SB, Ewertsen C, Carlsen JF et al. Ultrasound Elastography Is Useful for Evaluation of Liver Fibrosis in Children-A Systematic Review. J Pediatr Gastroenterol Nutr 2016; 63: 389-399
2. Kim JR, Suh CH, Yoon HM et al. The diagnostic performance of shear-wave elastography for liver fibrosis in children and adolescents: A systematic review and diagnostic meta-analysis. Eur Radiol 2018; 28: 1175-1186
3. de Franchis R, Fac BV. Expanding consensus in portal hypertension Report of the Baveno VI Consensus Workshop: Stratifying risk and individualizing care for portal hypertension. Journal of Hepatology 2015; 63: 743-752
4. Potter C, Hogan MJ, Henry-Kendjorsky K et al. Safety of pediatric percutaneous liver biopsy performed by interventional radiologists. J Pediatr Gastroenterol Nutr 2011; 53: 202-206
5. Dietrich CF, Bamber J, Berzigotti A et al. EFSUMB Guidelines and Recommendations on the Clinical Use of Liver Ultrasound Elastography, Update 2017 (Long Version). Ultraschall Med 2017; 38: e48
6. Bamber J, Cosgrove D, Dietrich CF et al. EFSUMB Guidelines and Recommendations on the Clinical Use of Ultrasound Elastography. Part 1: Basic Principles and Technology. Ultraschall Med 2013; 34: 169-184
7. Kettaneh A, Marcellin P, Douvin C et al. Features associated with success rate and performance of FibroScan measurements for the diagnosis of cirrhosis in HCV patients: a prospective study of 935 patients. J Hepatol 2007; 46: 628-634
8. Mulabecirovic A, Mjelle AB, Gilja OH et al. Liver elasticity in healthy individuals by two novel shear-wave elastography systems-Comparison by age, gender, BMI and number of measurements. PLoS One 2018; 13: e0203486
9. Fang C, Jaffer OS, Yusuf GT et al. Reducing the Number of Measurements in Liver Point Shear-Wave Elastography: Factors that Influence the Number and Reliability of Measurements in Assessment of Liver Fibrosis in Clinical Practice. Radiology 2018; 287: 844-852
10. Yoon JH, Lee JM, Han JK et al. Shear Wave Elastography for Liver Stiffness Measurement in Clinical Sonographic Examinations Evaluation of Intraobserver Reproducibility, Technical Failure, and Unreliable Stiffness Measurements. J Ultras Med 2014; 33: 437-447
11. Piscaglia F, Salvatorea V, Mulazzani L et al. Differences in liver stiffness values obtained with new ultrasound elastography machines and Fibroscan: A comparative study. Digest Liver Dis 2017; 49: 802-808
12. Ferraioli G, De Silvestri A, Lissandrin R et al. Evaluation of Inter-System Variability in Liver Stiffness Measurements. Ultraschall Med 2019; 40: 64-75
13. Mulabecirovic A, Mjelle AB, Gilja OH et al. Repeatability of shear wave elastography in liver fibrosis phantoms-Evaluation of five different systems. Plos One 2018; 13:
14. Sporea I, Gradinaru-Tascau O, Bota S et al. 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 2013; 15: 268-272
15. Thiele M, Madsen BS, Procopet B et al. Reliability Criteria for Liver Stiffness Measurements with Real-Time 2D Shear Wave Elastography in Different Clinical Scenarios of Chronic Liver Disease. Ultraschall Med 2017; 38: 648-654
16. Karlas T, Pfrepper C, Wiegand J et al. Acoustic radiation force impulse imaging (ARFI) for non-invasive detection of liver fibrosis: examination standards and evaluation of interlobe differences in healthy subjects and chronic liver disease. Scand J Gastroentero 2011; 46: 1458-1467
17. Procopet B, Berzigotti A, Abraldes JG et al. Real-time shear-wave elastography: Applicability, reliability and accuracy for clinically significant portal hypertension. Journal of Hepatology 2015; 62: 1068-1075

18. Zeng J, Liu GJ, Huang ZP et al. Diagnostic accuracy of two-dimensional shear wave elastography for the non-invasive staging of hepatic fibrosis in chronic hepatitis B: a cohort study with internal validation. Eur Radiol 2014; 24: 2572-2581
19. Cassinotto C, Lapuyade B, Mouries A et al. Non-invasive assessment of liver fibrosis with impulse elastography: comparison of Supersonic Shear Imaging with ARFI and FibroScan(R). J Hepatol 2014; 61: 550-557
20. Sporea I, Bota S, Jurchis A et al. Acoustic Radiation Force Impulse and Supersonic Shear Imaging Versus Transient Elastography for Liver Fibrosis Assessment. Ultrasound in Medicine and Biology 2013; 39: 1933-1941
21. Boursier J, Zarski JP, de Ledinghen V et al. Determination of reliability criteria for liver stiffness evaluation by transient elastography. Hepatology 2013; 57: 1182-1191
22. Mjelle AB, Mulabecirovic A, Havre RF et al. Normal Liver Stiffness Values in Children: A Comparison of Three Different Elastography Methods. J Pediatr Gastroenterol Nutr 2019; 68: 706-712