



## CLINICAL RESEARCH ARTICLE

# Repeatability of transient elastography in children

Marion Rowland<sup>1</sup>, Allison McGee<sup>2</sup>, Annemarie Broderick<sup>3</sup>, Brendan Drumm<sup>1</sup>, Lucy Connolly<sup>1</sup>, Leslie E. Daly<sup>4</sup>, Jennifer Drummond<sup>1</sup>, Emer Fitzpatrick<sup>5</sup>, Barry Linnane<sup>6</sup>, P. Aiden McCormick<sup>7</sup>, Paul McNally<sup>3,8</sup>, Louise Rainford<sup>2</sup>, Billy Bourke<sup>1,3</sup>, The Cystic Fibrosis Registry of Ireland<sup>9</sup> and The Cystic Fibrosis Liver Disease Research Group<sup>10</sup>

**BACKGROUND:** Poorly performing diagnostic tests can impact patient safety. Clinical investigations must have good precision and diagnostic accuracy before widespread use in clinical practice. Transient elastography (TE) measures liver stiffness, a surrogate marker of liver fibrosis in adults and children. Studies to evaluate its repeatability and reproducibility (precision) in children are limited. Our aim was to determine (i) the normal range of TE measurements and (ii) the repeatability and reproducibility of TE in healthy children.

**METHODS:** TE was performed in 257 healthy children, of whom 235 (91%, mean age 11.7 years, standard deviation (SD) 2.51, 107 were males (45.5%)) had two valid TE measurements performed, at least 24 h apart, by two operators under similar circumstances. High-quality TE images were obtained for each examination.

**RESULTS:** The normal range of TE was 2.88–6.52 kPa. The mean difference between paired measurements was 0.044 (SD 0.4). The 95% limits of agreement ranged from –0.8 to +0.76 kPa for repeat measurements. There was a difference of >1 kPa between measurements in 61/235 (25.9%) children. The lack of precision was similar across all age groups.

**CONCLUSIONS:** This study demonstrates that TE does not have acceptable precision in healthy children, because random measurement variation results in the lack of agreement between paired measurements.

*Pediatric Research* (2020) 88:587–592; <https://doi.org/10.1038/s41390-020-0916-4>

## IMPACT:

- The precision and diagnostic accuracy of a new technology must be determined before it is deployed in children in order to ensure that appropriate clinical decisions are made, and healthcare resources are not wasted.
- TE is widely used to diagnose liver disease in children without adequate evaluation of the precision (repeatability) of TE either in healthy children or children with liver disease.
- This study demonstrates that TE does not have adequate precision in children.
- This study was performed in accordance with methods previously published for children. Refinements to the test protocol, such as duration of fasting or probe size, will have to be evaluated for their impact on precision and accuracy before the test is deployed in research studies or clinical practice.

## INTRODUCTION

Transient elastography (TE) (FibroScan®, Echosens, Paris, France) is a rapid (5–10 min), non-invasive point-of-care technique, which measures liver stiffness.<sup>1,2</sup> Liver stiffness measurement (LSM) serves as a surrogate marker for the degree of liver fibrosis, reducing the need for invasive tests such as liver biopsy.<sup>1–3</sup> In adults an LSM cut-off of >13.0 kPa has been shown to have excellent diagnostic accuracy for liver cirrhosis.<sup>1,2</sup> Fraquelli et al.<sup>4</sup> evaluated the precision of TE in over 200 adults with liver disease. They reported high interclass correlation coefficients

(ICCs) for those with a Metavir score  $F \geq 3$ ; however, the ICC of 0.6 between repeated measures for milder forms of liver disease was poor.<sup>4</sup>

Cut-off values for milder degrees of fibrosis (Metavir stage  $\geq F2$ ) have not been well established and vary according to the underlying pathology.<sup>1,2</sup> Furthermore, a number of studies in adults have reported inconsistencies in repeated TE measurements in the absence of any change in the underlying pathology.<sup>5–7</sup> Discrepancies of 2–5 kPa have been described between paired measurements, which could result in misclassification of the stage

<sup>1</sup>School of Medicine, University College Dublin, Dublin, Ireland; <sup>2</sup>Radiography and Diagnostic Imaging, School of Medicine, University College Dublin, Dublin, Ireland; <sup>3</sup>Our Lady's Children's Hospital Crumlin, Dublin, Ireland; <sup>4</sup>School of Public Health, Physiotherapy and Sports Science, University College Dublin, Dublin, Ireland; <sup>5</sup>Paediatric Liver, GI and Nutrition Centre, Faculty of Life Sciences and Medicine, King's College London, London, UK; <sup>6</sup>Graduate Entry Medical School and Centre for Interventions in Infection, Inflammation and Immunity, Limerick University, Limerick, Ireland; <sup>7</sup>St Vincent's University Hospital, Dublin, Ireland; <sup>8</sup>Department of Pediatrics, School of Medicine, Royal College of Surgeons in Ireland, Dublin, Ireland; <sup>9</sup>Woodview House, Belfield, Dublin, Ireland and <sup>10</sup>Catherine McAuley Research Centre, Nelson Street, Dublin, Ireland  
Correspondence: Marion Rowland ([marion.rowland@ucd.ie](mailto:marion.rowland@ucd.ie))  
Deceased: Leslie E. Daly

Members of the Cystic Fibrosis Liver Disease Research Group are listed above the Acknowledgements.

These authors jointly supervised this work: Marion Rowland, Allison McGee

Received: 15 January 2019 Revised: 30 March 2020 Accepted: 1 April 2020

Published online: 1 May 2020

of fibrosis.<sup>3,6</sup> In addition, extrahepatic cholestasis has been shown to increase LSM irrespective of fibrosis stage.<sup>8</sup>

In children, TE is potentially an attractive non-invasive test to identify and monitor liver disease progression. The hallmark of any clinical investigation is that it is repeatable and reproducible over a range of clinical circumstances and conditions.<sup>9</sup> Therefore, careful evaluation of TE in children is required to ensure that measurements not only have good diagnostic accuracy (sensitivity and specificity) but also have precision in terms of repeatability and reproducibility.

While the diagnostic accuracy of TE has been examined in healthy children<sup>10–12</sup> and in children with a variety liver diseases,<sup>13–21</sup> few previous studies have examined the precision of TE in children.<sup>11,19</sup> Goldschmidt et al.,<sup>11</sup> using a subset of 28 children from a larger study population of healthy children, found that repeatability was good if TE measurements were performed in sequence using a marked point. However, repeatability was not satisfactory if examinations were carried out at different times by the same operator, or by different operators.

We planned to use TE to evaluate and monitor the progression of liver disease in a large cohort of children with Cystic Fibrosis (CF) (people with CF) as part of the national prospective study on the risk factors and outcome for Cystic Fibrosis Liver Disease (CFLD).<sup>22</sup> Before using TE as a research tool, we first had to confirm the precision of TE by examining repeatability and reproducibility in healthy children.

#### Aims

The aims of this study were to determine

- (i) the normal range of liver TE measurements in children,
- (ii) the repeatability and reproducibility of TE in healthy children.

## METHODS

### Research participants

Two sports clubs, under the auspices of the Gaelic Athletic Association, were approached and agreed to facilitate the study in healthy children. Eligible participants were informed of the study by sports club management and parents of interested volunteers contacted the research team who attended club facilities at agreed times. Following informed consent, children between 7 and 18 years of age were recruited over a 7-month period (March–September 2015).

Following consent and before performance of TE, each participant provided a short medical history, including medication usage, recent food intake, and had height and weight measured, as well as BMI calculated.<sup>23</sup> None of the 235 volunteers had a history of liver disease, or other gastrointestinal disease, and none were taking any regular medication. Children with a body mass index (BMI) greater than the 85th percentile were not included. Children under 7 years were not included because in our experience they find TE uncomfortable.

### Operator training

Both operators were experienced clinicians with a broad range of clinical, research and teaching expertise in diagnostic imaging and nursing. Prior to utilising the FibroScan® device investigators (A.McG., J.D.) attended small-group, 4-h training session provided by the manufacturer's training consultant. Based on their demonstrated ability to set up and correctly utilise the portable FibroScan® device, they were then certified by the manufacturer as competent to perform TE.

During preliminary TE preparation, operators (J.D. and A.McG.) encountered challenges in obtaining repeatable LSMs. Therefore, a complete re-evaluation of the operators' performance was

undertaken in conjunction with the manufacturer's training consultant. Sample time-motion mode (TM-mode) and amplitude-mode (A-mode) images were reviewed with the training consultant, seeking clarification and advice on improving quality. In order to avoid inadvertent inclusion of non-liver structures (right lung field, rib, vascular or biliary structures) within the region of interest, the operator must be able to clearly visualise speckle and line patterns associated with the TM- and A-mode images on the TE monitor.<sup>24</sup> To achieve this, operator seating position and room lighting were reviewed to ensure that only high-quality TM- and A-mode images were obtained and used to measure liver stiffness. Specifically, (i) correct seating position for the operator to ensure probe positioning perpendicular to the skin surface, in both planes, while allowing optimal viewing of the TE monitor, (ii) correct lighting, in accordance with American Association of Physicists in Medicine (AAPM) guidelines, with overhead room lights switched off during elastography scanning and a lamp was used to provide dimmed room lighting conditions.<sup>25</sup> In addition, the FibroScan® device was positioned to minimise reflections from windows or the lamp, which would have interfered with features in the displayed images.

### TE measurements

Using a portable FibroScan® machine, J.D. and A.McG. performed TE measurements on each participant at rest, on two separate occasions, at least 24 h apart, but within 3 weeks of the initial examination. An immediate repeat examination would not have been appropriate because persistent and easily observable skin indentation at the probe site could bias the second measurement.

Prior to positioning the patient, each operator ensured that seating lighting and monitor positioning were optimal, as outlined above. The participant was positioned in a comfortable supine position for the duration of the examination, with the right arm extended in maximal abduction with the hand behind the head so that there was easy access to position the probe perpendicular to the skin surface. The M probe of the FibroScan® machine was used according to the manufacturer's recommendations. The right lobe of the liver was accessed through an intercostal space at the level of the intersection of the mid-axillary line and a line extended laterally from the Xiphoid process. In practice, the transducer probe tip, covered with coupling gel, was positioned perpendicular to the skin surface in both axes, in the seventh right intercostal space, and moved or angled slightly anteriorly from the mid-axillary location, or moved superiorly or inferiorly to a higher or lower intercostal space in order to ensure optimal quality TM-mode images and A-mode graphs were generated, avoiding large intrahepatic vessels or other heterogeneous areas within the liver. When adequate images and graphs with no major artefacts/vessels were viewed in the TM- and A-modes, ten measurements were taken at the same location.

As per the manufacturer's guidelines, ten readings of liver stiffness were performed on each participant. The machine independently calculated the median (M), interquartile range (IQR), IQR/median (IQR/M) ratio and the number of valid measurements. A high IQR/M ratio implies a large distribution of valid LSMs and thus a higher risk of aberrant LSM median values. LSM accuracy using FibroScan® has been shown to decrease when the IQR/M ratio increases, and measurements with an IQR/M >30% have lower accuracy, while LSM measurements <10% have the highest accuracy, particularly with increasing measures of liver stiffness.<sup>26</sup> The manufacturer's criteria for a valid and reliable TE examination are a success rate of >60% (more than six valid readings) and an IQR/M ratio <30%. Participants with a success rate of <60% or an IQR/M ratio of >30% were automatically classified as a failed examination by the machine and excluded from the study.

To calculate the upper limit of normal (ULN), we included only those volunteers who had IQR/M ratios <25% on both TE

**Table 1.** Mean ( $\pm$ SD) values and upper and lower limits of normal for LSM for the first and second examination in healthy volunteers ( $n = 235$ ) according to age and gender.

	<i>n</i>	LSM (kPa)	IQR	IQR/median	Success rate	Lower limit normal	Upper limit normal
<b>Exam 1</b>							
Mean LSM	235	4.76 $\pm$ 0.85	0.59 $\pm$ 0.29	12.54 $\pm$ 5.48	96.78 $\pm$ 6.87	3.13	6.42
Age >12 years	105	4.99 $\pm$ 0.77*	0.63 $\pm$ 0.31	12.55 $\pm$ 5.51	97.65 $\pm$ 5.80	3.47	6.52
Age $\leq$ 12 years	130	4.57 $\pm$ 0.86	0.57 $\pm$ 0.28	12.52 $\pm$ 5.47	96.08 $\pm$ 7.58	2.93	6.23
Male	107	4.65 $\pm$ 0.91	0.59 $\pm$ 0.30	12.82 $\pm$ 5.29	97.41 $\pm$ 6.09	2.88	6.50
Female	128	4.85 $\pm$ 0.78	0.59 $\pm$ 0.29	12.29 $\pm$ 5.63	96.25 $\pm$ 7.44	3.38	6.32
<b>Exam 2</b>							
Mean LSM	235	4.67 $\pm$ 0.74	0.52 $\pm$ 0.23	11.39 $\pm$ 4.90	97.39 $\pm$ 6.25	3.42	6.04
Age >12 years	105	4.93 $\pm$ 0.69*	0.51 $\pm$ 0.19	10.48 $\pm$ 3.68**	98.00 $\pm$ 5.56	3.79	6.12
Age $\leq$ 12 years	130	4.47 $\pm$ 0.71	0.53 $\pm$ 0.25	12.12 $\pm$ 5.60	96.90 $\pm$ 6.73	3.23	5.83
Male	107	4.59 $\pm$ 0.77	0.54 $\pm$ 0.23	11.95 $\pm$ 4.99	96.45 $\pm$ 7.36	3.26	6.05
Female	128	4.75 $\pm$ 0.70	0.51 $\pm$ 0.22	10.91 $\pm$ 4.78	98.19 $\pm$ 5.02	3.57	6.01

LSM liver stiffness measurement.

\* $<0.001$ .

\*\* $<0.01$ .

measurements and a success rate of  $>80\%$  in order to include only measurements with a high degree of accuracy and reliability. Only 8/235 (3.4%) participants had two measurements with an IQR/M ratio of  $\leq 10\%$ .

#### Data analysis

Data were analysed with MedCalc (<https://www.medcalc.org>; 2016) (MedCalc Software, Ostend, Belgium). Data are presented as means with standard deviations (SDs) for continuous variables and the distribution of the sample was assessed with the Shapiro–Wilk test. As suggested by Bland and Altman, repeat measurements of liver stiffness by the same or two different operators were considered as different methods of measurement in this study.<sup>9,27</sup> To examine the variability in repeat measurements within the same subject, the SD of the difference between pairs of repeated measures was calculated.<sup>27</sup> Bland and Altman plots were used to demonstrate visually the degree of agreement between two observers. The differences between any pair of LSM were plotted against the mean of the measurements, indicating how large the disagreement was.<sup>9,27</sup> The repeatability coefficient was calculated according to the formula  $1.96 \times \text{SD}$ .<sup>28</sup> The repeatability coefficient is the difference, in kPa, that will be exceeded by 5% of pairs of measures on the same subject.

Correlation between measurements was analysed using Lin's concordance correlation coefficient (CCC).<sup>29</sup> Lin's CCC does not require the assumption of a normal distribution. The concordance coefficients were classified as poor ( $<0.90$ ), moderate (0.90–0.95), substantial (0.95–0.99) and excellent ( $>0.99$ ) (MedCalc Software, Ostend, Belgium 2016). The upper and lower limits of liver stiffness were calculated using the 97.5% quantile of the Student's *t*-distribution with  $n - 1$  degrees of freedom.

## RESULTS

Two hundred and fifty-seven healthy volunteer children were enrolled, and each had two TE examinations carried out at least 24 h apart. Data from 22 (8.6%) volunteers were classified as failed examinations because they did not meet the manufacturer's guidelines (success rate  $<60\%$ , IQR/M ratio  $>30\%$ ) for both TE examinations. Operator A conducted two examinations with 121 children, Operator B conducted two examinations with 71

children, while 43 children had one examination conducted by each operator.

The mean age of the 235 volunteers was 11.70 years (SD 2.51, range 7.01–17.12 years) and 107 (45.53%) were males. Girls were older (mean age 12.27 years, SD 2.53) than boys (mean age 11.02 years, SD 2.32,  $p < 0.001$ ).

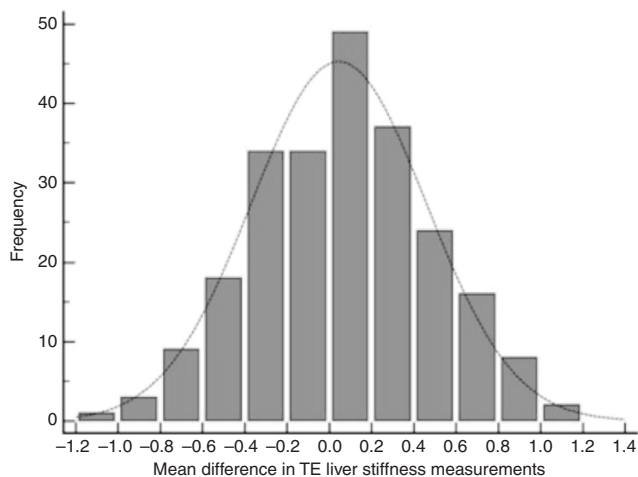
#### Normal range of liver stiffness in healthy children

The characteristics of TE measurements for Examination 1 and Examination 2 are outlined in Table 1. The mean LSM of Examination 1 (LSM 1) was 4.76 kPa, SD 0.85 kPa; and of Examination 2 (LSM 2) was 4.67 kPa, SD 0.74 kPa. Based on data from healthy children ( $n = 214$ ) who had two TE measurements fulfilling the following criteria: (i)  $\geq 80\%$  success rate and (ii) an IQR/M ratio of  $\leq 25\%$ , the range of normal LSM values in healthy children was established and ranged from 2.88 to 6.52 kPa (Table 1). As shown previously by others, the ULN was higher in children over 12 years of age compared to those under 12 years.<sup>10</sup> Gender did not significantly alter the ULN (Table 1).

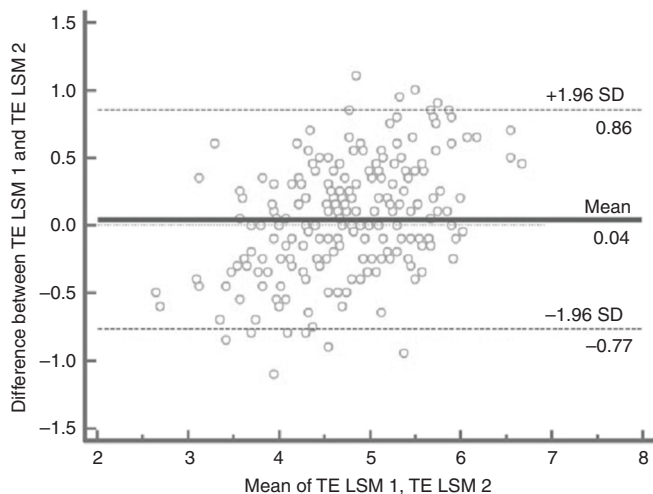
#### Agreement in healthy volunteers

The mean difference between paired measurements for liver stiffness for all healthy participants was  $-0.044$  kPa, SD 0.414 ( $p = \text{NS}$  paired *t* test). Figure 1 demonstrates that while the distribution of the paired differences for LSM 1 and LSM 2 followed a normal distribution (Shapiro–Wilk test  $W = 0.99$ ,  $p < 0.37$ ), there is a wide standard deviation of the differences. The 95% limits of agreement for repeated TE measurements ranged from  $-0.85$  to  $+0.76$  kPa in 235 healthy volunteers (Fig. 2). There was a wide scatter of data points with 15/235 (6.4%) lying on or outside the 95% limits of agreement. There was a difference of  $\geq 1$  kPa between the two examinations in 61/235 (25.9%) participants. The repeatability coefficient was 0.811 kPa.

When a single operator carried out both measurements (Operator A,  $n = 121$ ; Operator B,  $n = 71$ ), the difference between the upper and lower limits of agreement for Operator A was 1.51 kPa ( $-0.79$  to  $+0.718$ ), (Supplemental Fig. S1) with a repeatability coefficient of 0.75 kPa, and for Operator B was 1.80 kPa ( $-0.87$  to  $0.93$ ) repeatability coefficient 0.89 kPa. The level of disagreement (1.66 kPa) was similar when two different operators carried out the initial and repeat measurement ( $n = 43$ ; Supplementary Fig. S2) with a repeatability coefficient of 0.83 kPa.



**Fig. 1 Histogram of Paired Differences in TE measurements for 235 healthy children.** Histogram of the mean difference between liver stiffness measurement 1 (LSM 1) and liver stiffness measurement 2 (LSM 2) in  $n = 235$  healthy children demonstrating the wide standard deviation (0.41) of the mean difference between the paired measurements.



**Fig. 2 Bland and Altman Plot of the agreement between two TE measurements in 235 healthy children.** Bland and Altman plot that demonstrates the 95% limits of agreement ( $-0.77$  to  $+0.86$  kPa), between two liver stiffness measurements in 235 healthy children.

#### Within-subject variability

We used Lin's CCC to examine within-subject variability and found that the concordance between two measurements in the same individual was poor, 0.85 (95% confidence interval (CI): 0.82–0.88;  $n = 235$ ). When the same operator performed both examinations, within-subject variability was also high with poor CCCs, Operator 1: 0.86 (95% CI: 0.81–0.89); Operator 2: 0.85 (95% CI: 0.75–0.91).

#### The effect of age on precision of TE

There was no difference in the mean and standard deviation of paired differences in LSM for children under 14 years of age ( $n = 190$ , mean difference = 0.045, SD 0.41) compared to children over 14 years of age ( $n = 45$ , mean difference = 0.038, SD 0.41,  $p = \text{NS}$ ). Neither was there a difference in paired LSM measurements for children under 10 years of age ( $n = 64$ , mean difference =  $-0.004$ , SD 0.46) compared to children over 10 years of age ( $n = 171$ , mean difference = 0.06, SD 0.39,  $p = \text{NS}$ ).

## DISCUSSION

TE is now widely used in the evaluation of children with liver disease.<sup>13–20</sup> In this study, our findings on the normal range of TE values in healthy children are consistent with those reported in other studies.<sup>10–12</sup> However, we found that TE lacks acceptable precision as a diagnostic instrument for use in children.

Poorly performing diagnostic tests can negatively impact on patient safety and waste scarce medical resources.<sup>30</sup> The terminology to describe the performance of a clinical instrument can be confusing, and the literature on diagnostic accuracy studies abounds with a range of different terms to describe the two important facets of any diagnostic test or instrument.<sup>28,31,32</sup> When a new instrument or test is evaluated, it is important to determine both its precision (that on average two or more measurements taken over a short period of time will be the same) and diagnostic accuracy (that the instrument can clearly distinguish those with disease from the healthy population).<sup>33</sup> Measures of precision include reliability, repeatability, reproducibility or agreement, while test accuracy is reported as sensitivity and specificity or predictive value.<sup>33</sup> The prevalence of the disease determines the accuracy of the test, and therefore diagnostic accuracy in children should not be inferred from adult studies.<sup>1,33</sup> In the case of TE, there must be evidence demonstrating clinically acceptable levels of both precision and diagnostic accuracy, before TE is widely deployed as a non-invasive test to diagnose or monitor liver disease in children.<sup>1,9,27,30</sup>

In this study, we assessed the precision of TE in healthy children, that is, how variable repeated TE measurements are when made by the same observer or different observers on the same child. We also examined the repeatability of TE in different age groups. We assumed, when planning our study, that there would be negligible differences between two examinations a short time apart. However, we found that the precision of TE in children was poor. Variability occurred irrespective of the age of the child or whether the test was performed by the same operator or different operators.

Bland and Altman developed the limits of agreement method to examine the differences between measurements made by two methods or two observers.<sup>9,27</sup> Central to the method is examining the mean and standard deviation of the distribution of paired differences.<sup>27</sup> A wide standard deviation demonstrates lack of agreement between observers or tests.<sup>9,27</sup> This study demonstrates that while the distribution of paired differences follows a normal distribution, the standard deviation is wide (Fig. 1), signifying that agreement between repeated TE measurements is inadequate.<sup>27</sup> For 95% of healthy children, a given measurement of liver stiffness could range between 0.85 kPa less than or 0.76 kPa greater than a second measurement within a short period of time (Fig. 2). Over one-third of healthy children had an actual difference of 0.8 kPa between the first and second measurement, while 61/235 (25.9%) had a difference of  $>1$  kPa between measurements. While these differences appear inconsequential, they could result in a change in liver disease classification based on TE measurements without any change in the underlying pathology. Adult studies have also reported discrepancies in repeated TE measurements, which could result in a change of liver disease classification.<sup>3,5,7</sup>

In clinical practice measurement variation is inevitable, but the degree of variation that can be deemed acceptable is determined by what constitutes a clinically important difference between measurements. Initial investigative studies of diagnostic accuracy should include an evaluation of the precision of the test. However, as noted by Harris and Smith<sup>32</sup> and by Watson and Petrie<sup>34</sup> procedures to assess reliability and measure agreement are often overlooked. In the first study of TE in adults, Sandrin et al.<sup>35</sup> provided precision data in only 15/91 (16%) participants. From the data presented, it is not possible to determine if TE had clinically acceptable levels of precision. Subsequently, Fraquelli et al.<sup>4</sup>

examined the repeatability of TE in adults with biopsy-proven liver disease. They reported that while there was good agreement between repeat examinations in those with cirrhosis, TE precision was poor in those with milder forms of liver disease.<sup>4</sup>

There are only two paediatric studies<sup>11,19</sup> prior to ours examining the precision of TE in children, and both highlight the inherent difficulties of getting good agreement between repeated measures.<sup>11,19</sup> Goldschmidt et al.<sup>11</sup> examined the repeatability of TE in 28 of 504 (5.5%) children and reported that agreement was good if TE measurements were performed in sequence using a marked point. However, repeatability was not satisfactory if examiners conducted the TE examination independently, which is reflective of clinical practice. Nobili et al.,<sup>19</sup> in a study of children with non-alcoholic fatty liver disease, reported that they used ultrasound guidance, to determine optimal probe position and obtain good measures of interobserver agreement. However, even with ultrasound guidance the authors report a wide CI at the 90% level for the ICC (ICC 0.96, 90% CI: 0.92–0.97), indicating poor precision. Taken together with the data in this study, it suggests that the precision of TE in children is unacceptable as a non-invasive marker of liver disease.

We have considered the potential impact of a number of factors, including probe size, participant's age and operator capacity, which may explain our results. LSM measurements are not comparable when different probe sizes are used,<sup>11,36</sup> and therefore we used the M probe for all children in this study. Although the M probe was not recommended by the manufacturer for children under 14 years of age, it has been widely used in paediatric studies both before and after the development of the smaller paediatric probes.<sup>13,14,17,19–21,37</sup> Goldschmidt et al.<sup>11</sup> report that optimal evaluation of liver stiffness in children requires the use of the largest size probe that achieves a satisfactory LSM output. They report that 7/42 children (16.3%) aged 10.3–17.2 years would have been classified as "significant fibrosis" (cut-off >7.0 kPa) if the smaller S2 probe had been used.<sup>11</sup> Given that over 90% of participants in this study had >80% success rate for both examinations with IQR/M of <25%, it is unlikely that probe size can explain the lack of precision reported in this study.

To further evaluate the impact of probe size, we examined the effect of age on the standard deviation of the paired differences in LSM and demonstrate that even in children over the age of 14 years, for whom the M probe is recommended by the manufacturer, the SD of the paired differences is wide supporting a lack of precision of TE regardless of age.

Operator capacity is unlikely to explain our results. The manufacturer certified the training of our operators. The range of normal TE values documented in children in this study is very similar to that reported in other paediatric studies.<sup>10–12</sup> Our rate of failed examinations (success rate <60% and IQR/M <30%) was 8.6% (22/257), which is less than the failure reported by others.<sup>11,15</sup> Furthermore, we achieved an 80% success rate for both examinations in 91% (214/235) of participants, which suggests that operator capacity was very good.

Children were not included in this study if they had a substantial meal within 2 h of the TE examination. However, strict fasting was not a requirement for participation. Food intake prior to TE is an evolving area and a 3-h fast is now a requirement for TE examinations based on data from a number of adult studies.<sup>38,39</sup> The evidence for a prolonged fast in children is ambiguous. TE examinations carried out 30 min after lunch have been shown to increase LSM values in healthy children.<sup>11,12</sup> In children with biopsy-proven liver disease, Lee et al.<sup>15</sup> reported that there were no differences in LSM values between children who fasted overnight ( $n = 40$ ) and those who did not fast. Metavir score was the only determinant of differences in LSM.<sup>15</sup> While the requirement for a 3 h fast is now considered necessary, prolonged fasting needs careful evaluation on the accuracy and precision of TE in children.

This study has a number of weaknesses. It was not initially designed as a precision or diagnostic accuracy study of TE, but rather to demonstrate that TE measurements could monitor change in the liver disease status of children with CF. It includes only healthy volunteers and relies on clinical history as the gold standard for the absence of liver disease. Liver biopsy is the gold standard for the diagnosis of liver disease, but liver biopsy is not a clinically appropriate investigation in adults or children unless there is a high suspicion of liver disease.

It is recommended that studies of diagnostic accuracy should follow Standards for Reporting of Diagnostic Accuracy (STARD)<sup>40</sup> reporting guidelines. While this study was not a diagnostic accuracy study of the sensitivity and specificity of TE against a gold standard, rather an examination of agreement between repeated measurements, we report our data in so far as is possible in line with the principles of transparency as outlined in STARD.

The strength of this study was the early identification of issues with repeated TE measurements, carefully considered approaches to training and optimisation of our protocols and the inclusion of over 200 healthy children to evaluate the repeatability of TE.

As part of a national prospective study on risk factors and outcome for CFLD, we also wanted to determine if TE had the precision required to facilitate the early diagnosis of CFLD and monitor disease progression. However, recruitment of participants with CF was stopped because of the lack of precision of TE demonstrated in healthy children.

At the time this study was stopped, we had enrolled 137 participants with CF, of whom 128 had a valid TE examination; 20 (15.6%) had liver disease (CFLD) with clinical or radiological evidence of portal hypertension, 43 (33.6%) had nonspecific changes on ultrasound or biochemical indices (Non-specific Cystic Fibrosis Liver Disease (NSCFLD)), while 65 (50.8%) had no evidence of liver disease (No liver Disease (NoLD)) as outlined in Supplemental Data (online).

Sixty-six participants with CF had two examinations on separate days, of whom 11 had CFLD, 26 had NSCFLD and 29 had NoLD. The mean difference between paired measurements for the 66 participants with CF was  $-0.41$ , SD 3.22 kPa (paired  $t$  test NS). Supplemental Figure S3 (online) shows the wide standard deviation between paired measurements that was not normally distributed. Using a Bland and Altman plot, the limits of agreement between the two measurements of liver stiffness ranged from  $-5.0$  to 6.1 kPa (Supplementary Fig. S4 (online)). In those with CFLD ( $n = 11$ ), the mean difference between paired measurements was 0.02 kPa with a SD of 5.05 and the limits of agreement ranged from  $-9.9$  to 9.9 kPa (Supplemental Fig. S5 (online)). The repeatability coefficient for those with CFLD was 9.8 kPa. Concordance between two measures was poor, regardless of the CF liver disease status of participants. In those with NoLD, the CCC was 0.86 (95% CI: 0.72–0.93), while in those with CFLD, it was 0.82 (95% CI: 0.48–0.95).

The disagreement demonstrated between two TE measurements over a short time in participants with or without liver disease could change the classification of their disease without any change in the underlying pathology. These findings on the lack of precision of TE in children with liver disease are consistent with our findings in healthy children that we report in this study.

## CONCLUSION

Adequate evaluation of any new technology is essential to ensure that appropriate clinical decisions lead to optimal recommendations for treatment.<sup>30</sup> This study demonstrates that TE does not have acceptable precision in children, because random measurement variation results in the lack of agreement between paired examinations. Further development is required to optimise the precision of TE in children in order to ensure patient safety, appropriate clinical decision-making and the optimal use of healthcare resources.

## CYSTIC FIBROSIS LIVER DISEASE RESEARCH GROUP

B. Casserly<sup>11</sup>, D. Cox<sup>3</sup>, A. Das<sup>12</sup>, C. Gallagher<sup>7</sup>, P. Greally<sup>13</sup>, C. Gunaratnam<sup>14</sup>, F. Healy<sup>15</sup>, M. Herzig<sup>16</sup>, S. Javadpour<sup>3</sup>, G. Lean<sup>13</sup>, N. G. McElvaney<sup>14</sup>, E. McKone<sup>7</sup>, D. Mullane<sup>17</sup>, M. Ni Chroinin<sup>17</sup>, M. O' Mahony<sup>16</sup>, M. O' Neil<sup>18</sup>, B. Plant<sup>17</sup>, M. Rogan<sup>12</sup> and C. Short<sup>17</sup>

<sup>11</sup>University Hospital Limerick, Limerick, Ireland; <sup>12</sup>Waterford Regional Hospital, Waterford, Ireland; <sup>13</sup>National Children's Hospital, Tallaght, Dublin, Ireland; <sup>14</sup>Beaumont Hospital, Dublin, Ireland; <sup>15</sup>Children's University Hospital, Temple Street, Dublin, Ireland; <sup>16</sup>Galway University Hospital, Galway, Ireland; <sup>17</sup>Cork University Hospital, Cork, Ireland and <sup>18</sup>Castlebar General Hospital, Mayo, Ireland

## ACKNOWLEDGEMENTS

We thank all those who participated in this study, in particular our healthy volunteers without whom this study would not have been possible. This research was funded by The Health Research Board (HRA-PHS-662). The Health Research Board had no role in the study. There was no industry support for this study. The corresponding author had full access to all the data and has final responsibility for the decision to submit for publication.

## AUTHOR CONTRIBUTIONS

Study concept: M.R., B.B. and P.A.McC. Study design: M.R. and L.E.D. Acquisition of data: A.McG., J.D. and L.R. Data analysis: A.McG., L.E.D. and M.R. Data interpretation: B. L., L.R., E.F., B.B., J.D., P.McN., A.McG. and M.R. First draft manuscript: M.R. and A.McG. Critical revision of manuscript: B.B., B.D., A.McG. and M.R. Final appraisal of manuscript: B.B., B.D., P.A.McC., B.L., E.F., L.R., A.B., J.D., L.C., P.McN., Cystic Fibrosis Registry of Ireland and Cystic Fibrosis Liver Disease Research Group

## ADDITIONAL INFORMATION

The online version of this article (<https://doi.org/10.1038/s41390-020-0916-4>) contains supplementary material, which is available to authorised users.

**Competing interests:** The authors declare no competing interests.

**Ethical approval:** The study was approved by the Ethics Committees of Our Lady's Children's Hospital Crumlin (for both healthy volunteers and children with cystic fibrosis) and by the National Children's Hospital Tallaght, Children's University Hospital and St Vincent's University Hospital Dublin for participants with cystic fibrosis. Written informed consent was obtained from parents or guardians of children under 18 years of age for their child's participation in the study with assent from the children. Participants over 18 years of age provided written informed consent for the study.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## REFERENCES

- Friedrich-Rust, M., Poynard, T. & Castera, L. Critical comparison of elastography methods to assess chronic liver disease. *Nat. Rev. Gastroenterol. Hepatol.* **13**, 402–411 (2016).
- Tsochatzis, E. A. et al. Elastography for the diagnosis of severity of fibrosis in chronic liver disease: a meta-analysis of diagnostic accuracy. *J. Hepatol.* **54**, 650–659 (2011).
- Vuppalanchi, R. & Sanyal, A. J. Myths and mysteries about staging hepatic fibrosis by fibroscan. *Clin. Gastroenterol. Hepatol.* **13**, 780–782 (2015).
- Fraquelli, M. et al. Reproducibility of transient elastography in the evaluation of liver fibrosis in patients with chronic liver disease. *Gut* **56**, 968–973 (2007).
- Nascimbeni, F. et al. Significant variations in elastometry measurements made within short-term in patients with chronic liver diseases. *Clin. Gastroenterol. Hepatol.* **13**, 763–771 e1 (2015).
- Vignier, N. et al. Reproducibility of liver stiffness measurements in hepatitis C virus (HCV)-infected patients in Egypt. *J. Viral Hepat.* **18**, e358–e365 (2011).
- Roca, B., Resino, E., Torres, V., Herrero, E. & Penades, M. Interobserver discrepancy in liver fibrosis using transient elastography. *J. Viral Hepat.* **19**, 711–715 (2012).
- Millonig, G. et al. Extrahepatic cholestasis increases liver stiffness (FibroScan) irrespective of fibrosis. *Hepatology* **48**, 1718–1723 (2008).
- Bland, J. M. & Altman, D. G. Measuring agreement in method comparison studies. *Stat. Methods Med. Res.* **8**, 135–160 (1999).

- Engelmann, G. et al. Feasibility study and control values of transient elastography in healthy children. *Eur. J. Pediatr.* **171**, 353–360 (2012).
- Goldschmidt, I. et al. Application and limitations of transient liver elastography in children. *J. Pediatr. Gastroenterol. Nutr.* **57**, 109–113 (2013).
- Lewindon, P. J. et al. Transient liver elastography in unsedated control children: impact of age and intercurrent illness. *J. Paediatr. Child Health* **52**, 637–642 (2016).
- Fitzpatrick, E., Quaglia, A., Vimalaevan, S., Basso, M. S. & Dhawan, A. Transient elastography is a useful noninvasive tool for the evaluation of fibrosis in paediatric chronic liver disease. *J. Pediatr. Gastroenterol. Nutr.* **56**, 72–76 (2013).
- Lee, C. K. et al. Serum biomarkers and transient elastography as predictors of advanced liver fibrosis in a United States cohort: the Boston children's hospital experience. *J. Pediatr.* **163**, 1058–1064 e2 (2013).
- Lee, C. K. et al. Validation of transient elastography cut points to assess advanced liver fibrosis in children and young adults: The Boston Children's Hospital Experience. *J. Pediatr.* **198**, 84–89 e2 (2018).
- Malbrunot-Wagner, A. C. et al. Transient elastography and portal hypertension in pediatric patients with cystic fibrosis. *J. Cyst. Fibros.* **10**, 338–342 (2011).
- Menten, R. et al. Transient elastography in patients with cystic fibrosis. *Pediatr. Radiol.* **40**, 1231–1235 (2010).
- Friedrich-Rust, M. et al. Non-invasive measurement of liver and pancreas fibrosis in patients with cystic fibrosis. *J. Cyst. Fibros.* **12**, 431–439 (2013).
- Nobili, V. et al. Accuracy and reproducibility of transient elastography for the diagnosis of fibrosis in pediatric nonalcoholic steatohepatitis. *Hepatology* **48**, 442–448 (2008).
- Witters, P. et al. Non-invasive liver elastography (Fibroscan) for detection of cystic fibrosis-associated liver disease. *J. Cyst. Fibros.* **8**, 392–399 (2009).
- Behairy Bel, S., Sira, M. M., Zalata, K. R., Salama el, S. E. & Abd-Allah, M. A. Transient elastography compared to liver biopsy and morphometry for predicting fibrosis in pediatric chronic liver disease: Does etiology matter? *World J. Gastroenterol.* **22**, 4238–4249 (2016).
- Rowland, M. Liver disease in cystic fibrosis—a clinical challenge in the era of improved outcomes. *J. Cyst. Fibros.* **18**(Suppl. 1), S143–S144 (2019).
- HSE. *UK-WHO-Ireland Growth Charts* <https://www.hse.ie/eng/health/child/growthmonitoring/> (2013).
- Tang, A., Cloutier, G., Szevevényi, N. M. & Sirlin, C. B. Ultrasound elastography and MR elastography for assessing liver fibrosis: Part 1, principles and techniques. *Am. J. Roentgenol.* **205**, 22–32 (2015).
- Moore, S. C., Munnings, C. R., Brettell, D. S. & Evans, J. A. Assessment of ultrasound monitor image display performance. *Ultrasound Med. Biol.* **37**, 971–979 (2011).
- Boursier, J. et al. Determination of reliability criteria for liver stiffness evaluation by transient elastography. *Hepatology* **57**, 1182–1191 (2013).
- Altman, D. G. Assessing new methods of clinical measurement. *Br. J. Gen. Pract.* **59**, 399–400 (2009).
- McAlinden, C., Khadka, J. & Pesudovs, K. Precision (repeatability and reproducibility) studies and sample-size calculation. *J. Cataract Refract. Surg.* **41**, 2598–2604 (2015).
- Barnhart, H. X., Lakhnygina, Y., Kosinski, A. S. & Haber, M. Comparison of concordance correlation coefficient and coefficient of individual agreement in assessing agreement. *J. Biopharm. Stat.* **17**, 721–738 (2007).
- McGrath, T. A. et al. Overinterpretation of research findings: evidence of “Spin” in systematic reviews of diagnostic accuracy studies. *Clin. Chem.* **63**, 1353–1362 (2017).
- Lord, S. J., Staub, L. P., Bossuyt, P. M. & Irwig, L. M. Target practice: choosing target conditions for test accuracy studies that are relevant to clinical practice. *BMJ* **343**, d4684 (2011).
- Harris, E. F. & Smith, R. N. Accounting for measurement error: a critical but often overlooked process. *Arch. Oral. Biol.* **54**(Suppl. 1), S107–S117 (2009).
- Anvari, A., Halpern, E. F. & Samir, A. E. Statistics 101 for radiologists. *RadioGraphics* **35**, 1789–1801 (2015).
- Watson, P. F. & Petrie, A. Method agreement analysis: a review of correct methodology. *Theriogenology* **73**, 1167–1179 (2010).
- Sandrin, L. et al. Transient elastography: a new noninvasive method for assessment of hepatic fibrosis. *Ultrasound Med. Biol.* **29**, 1705–1713 (2003).
- Kim, S. et al. Points to be considered when applying FibroScan S probe in children with biliary atresia. *J. Pediatr. Gastroenterol. Nutr.* **59**, 624–628 (2014).
- Hwang, J. Y. et al. Diagnostic performance of transient elastography for liver fibrosis in children: a systematic review and meta-analysis. *Am. J. Roentgenol.* **211**, W257–W266 (2018).
- Arena, U. et al. Liver stiffness is influenced by a standardized meal in patients with chronic hepatitis C virus at different stages of fibrotic evolution. *Hepatology* **58**, 65–72 (2013).
- Berzigotti, A. et al. Effect of meal ingestion on liver stiffness in patients with cirrhosis and portal hypertension. *PLoS ONE* **8**, e58742 (2013).
- Bossuyt, P. M. et al. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. *BMJ* **351**, h5527 (2015).