Shear Wave Elastography in the assessment of Liver Fibrosis

Dr David Nadebaum

Master of Philosophy Thesis
The University of Melbourne

Department of Gastroenterology & Hepatology and Department of Radiology
Royal Melbourne Hospital

Abstract

The accurate quantification of liver fibrosis is essential to the prognostication and clinical management of patients with chronic liver disease (CLD). Whilst liver biopsy remains the gold standard for fibrosis assessment, it has a number of limitations which have seen its use become increasingly substituted by non-invasive techniques. Ultrasound shear wave elastography (SWE) includes some of the most widely used non-invasive technologies in clinical practice. This work evaluates two ultrasound SWE devices which are in differing stages of clinical development and use; the first being a well-validated point SWE technique from Siemens called Acoustic Radiation Force Impulse elastography or 'ARFI' and the second a new 2D-SWE platform by Toshiba. The differing study aims for the two technologies were assessed in separate patient cohorts. Hence the thesis is divided in two.

ARFI (Siemens)

Background

Acoustic Radiation Force Impulse elastography or 'ARFI' is a point shear wave elastography (SWE) technique that is in broad clinical use for the quantification of liver fibrosis. Whilst well validated, questions remain for a number of areas of ARFI performance. This includes the magnitude and likely mechanism of obesity's impact on ARFI performance, the impact of hepatosteatosis on ARFI reliability and whether ARFI performance is dependent on operator experience. There is also conflicting information as to whether ARFI liver stiffness measurements (LSMs) correlate with cirrhosis severity and the presence of cirrhotic complications. Finally, clinicians have limited facility to gauge the validity of obtained ARFI measurements beyond the IQR/Median criteria. An additional study aim was therefore to develop new strategies to aid ARFI reliability assessment; specifically whether inter-operator disagreement predicts the presence of unreliable ARFI measurements.

Method

ARFI performance was assessed amongst a cohort of 943 patients with diffuse CLD of mixed aetiology, who had ARFI LSMs taken as part of clinical fibrosis assessment. Patients were scanned independently by either two or three operators, with ARFI results analysed in the context of patient demographic and CLD information obtained from medical records. Anthropometric measures including body mass index (BMI) was recorded at the

time of scanning, and the distance from the skin surface to liver capsule (SLD) was measured from ARFI screenshots as a marker of central adiposity. The cumulative number of scans completed by individual operators and the institution overall was recorded. Assessed performance measures included IQR/Median and inter-operator agreement. ARFI accuracy was also assessed amongst a subcohort of 55 patients who had undergone a liver biopsy within 6 months of ARFI.

The performance of ARFI in assessing cirrhosis severity was assessed amongst a further subcohort of 186 patients with clinically diagnosed cirrhosis. The presence of cirrhotic complications was determined retrospectively from medical records and endoscopy reports. Prognostic indices including Child Pugh and Model for End stage Liver Disease (MELD) scores were calculated using bloods tests where available.

Results

ARFI showed modest accuracy in assessing liver fibrosis, demonstrating an AUROC of 0.67, 0.76 and 0.70 at discriminating the F01/ F2, F2/ F3 and F3/F4 cut-offs, respectively. ARFI showed good sensitivity (80.0 – 88.9%) and NPV (70.6 – 95.3%), but relatively poor specificity (42.9 – 66.3%) and PPV (27.9 – 56.2%) at the three cut-offs.

Body habitus, particularly skin-to-liver capsule distance or 'SLD', was found to be the primary determinant of ARFI performance in multi-regression analyses. SLD had the strongest relationship with ARFI accuracy (R^2 = 0.543) followed by necroinflammatory change (R^2 = 0.167), whilst all other patient factors, including hepatosteatosis, failed to show an independent association. Patients with a SLD >2.5cm (indicating significant central adiposity) showed particularly poor ARFI performance and was associated with higher IQR/Median ratios (median = 0.363 vs. 0.187, p<0.001), greater deviation between operators (29.8% vs. 15.9%, p<0.001) and poorer correlation with biopsy (rho = -0.242 vs. 0.493) than those with a SLD ≤2.5cm. Individual operator experience showed a weak relationship with ARFI performance, with operators of <25 scans experience having similar median IQR/Median ratios (0.170 vs. 0.165, p=0.13), slightly greater deviation between operators (14.3% vs. 11.06%, p=0.014) and greater deviation from the biopsy reference range (mean deviation = 0.588 vs. 0.279m/s, p=0.004) than more experienced colleagues. There also appeared to be a similarly weak association between overall institutional experience and ARFI performance, with reliability being slightly reduced amongst the first 150 scans performed in the institution.

In patients in whom both operators had concordant F score results, ARFI LSM showed greater correlation with biopsy (rho = 0.392) than in cases of inter-operator disagreement (rho = 0.010). When scanned by three operators, patients with three-way operator agreement showed even stronger correlation with histopathology (rho = 0.571).

Amongst cirrhotic patients, ARFI showed a moderately strong correlation with prognostic scores of liver function, including both MELD score (rho = 0.342, $p<0.001$) and Child-Pugh Score (rho = 0.363, $p<0.001$). ARFI LSMs showed modest accuracy in predicting the presence of ascites (AUROC = 0.58), encephalopathy (AUROC = 0.60) and oesophageal varices (AUROC = 0.69).

Conclusion

ARFI showed moderate performance in quantifying liver fibrosis in a clinical Australian setting. The technology's strength appears to be in the exclusion of liver fibrosis, however the tool is prone to false positive results. Body habitus was found to be the primary determinant of ARFI performance, with necroinflammatory change and operator experience showing a weaker impact on scan reliability. Central adiposity, as indicated by SLD, showed a particularly strong relationship with ARFI performance and the routine measurement and reporting of SLD should be considered to help clinicians gauge the reliability of ARFI results. Scanning patients with multiple independent operators also showed value as a reliability indicator, with inter-operator discordance being a predictor of poor ARFI performance.

2D-SWE (Toshiba)

Background

The second technology assessed is a new 2D-SWE platform from Toshiba, which has a number of technical innovations and theoretical advantages over Siemens' ARFI system. The technology is in the early clinical phases of testing and therefore data on this new technique remains limited. Our study aim was therefore to evaluate specific technical parameters to help assist in the formation of acquisition guidelines. This included assessing the measurement variability of Toshiba 2D-SWE (i.e. IQR/Median), the number of measurements required per patient to yield a precise LSM estimate and whether the

uniformity of shear wave velocities within the measurement ROI (i.e. ROI SD/Speed ratio) could be used to assess the reliability of individual 2D-SWE measurements.

Method

2D-SWE was assessed amongst fifty-five patients with mixed aetiology CLD using the Toshiba Aplio 500 ultrasound system. Ten measurements were obtained per patient by an operator blinded to all preceding readings. Measurement variability (i.e. IQR/Median) and the number of measurements required per patient to achieve a LSM estimate within 5% of the existing method using 10 samples was assessed. Results were analysed against scan and clinical information including CLD aetiology, BMI, SLD, presence and severity of hepatosteatosis and measurement depth within the liver. The ratio of the standard deviation of shear wave velocities within the measurement ROI to overall shear wave velocity (i.e. ROI SD/Speed) was calculated for each individual measurement, and its relationship with measurement consistency (i.e. deviation of the measurement from the set's median) was assessed.

Results

The median IQR/Median ratio for 2D-SWE was 0.131 (q1-q3: 0.089–0.174). Five readings provided an approximation within 0.11m/s or 4.2% of the median velocity of ten measurements. Factors associated with increased measurement variability included increasing BMI (rho=0.388, p=0.003), SLD (rho=0.426, p=0.002) and measurements taken within 1.5cm of the liver capsule ($p<0.001$). Measurements with heterogeneous shear wave profiles (indicated by a ROI SD/Speed >0.15) showed greater deviation from the set's median velocity than those with a ROI SD/Speed ≤0.15 (0.421 vs. 0.219 m/s, p=0.0001).

Conclusion

2D-SWE showed low overall measurement variability, with a minimum of five readings providing equivalent precision to the existing method using 10 samples. Obesity (i.e. BMI>30kg/m^2), increasing abdominal wall thickness (i.e. SLD), sub-capsular measurements and a ROI SD/Speed >0.15 were all associated with increased measurement variability. ROI SD/Speed warrants further evaluation as a quality assessment metric, as it may allow objective operator assessment of individual 2D-SWE measurement reliability in real-time.

Declaration

I declare that all of the work encompassed in this thesis is my own, unless otherwise acknowledged / referenced.

Acknowledgements

Special acknowledgement is required for my two primary supervisors, A/Prof Amanda Nicoll and Prof Robert Gibson, for their guidance, wisdom and patience. I am extremely grateful for their ongoing support, without which this work would not be possible.

Thank is also required for Ms Alexandra Gorelik and Prof Danny Liew (Melbourne EpiCentre) for their expert guidance regarding study design and statistical support.

Acknowledgement is required to Dr Jessica Howell, who helped organise ethics approval for the ARFI analyses prior to my involvement in the study.

Dr Michael Christie also requires acknowledgment for his assistance in reviewing the liver biopsy specimens analysed in the study.

Data from this project have been used by colleagues, Dr Shu Su, Dr Nicole Harris and Dr Kidd Leong, in the preparation of journal manuscripts. Some discussion regarding the mechanism underlying ARFI's poor correlation with cirrhosis severity was drawn from work published by Dr Leong. The data collection, statistical analyses and remainder of the interpretation and write up has been performed independently, without their input.

Table of Contents

Abbreviations

2D-SWE	Two-dimensional shear wave elastography
95%CI	95% Confidence Interval
A score	Metavir Activity score (i.e. for necroinflammation)
AIH	Autoimmune hepatitis
ALT	Alanine aminotransferase
APRI	AST to Platelet Ratio Index
AUROC	Area Under the Receiver Operator Curve
ARFI	Acoustic Radiation Force Impulse
AST	Aspartate aminotrasferase
BMI	Body Mass Index
CLD	Chronic Liver Disease
cm	centimetres
CT	Computed Tomography
DILI	Drug Induced Liver Injury
ECM	Extra-Cellular matrix
EFSUMB	European Federation of Societies of Ultrasound in Medicine and Biology
F score	Fibrosis score
F0	No fibrosis
F1	Minimal fibrosis
F2	Significant fibrosis
F3	Severe fibrosis
F4	Cirrhosis
FBE	Full Blood Examination

FIB-4	Fibrosis 4 Score
FOV	Field of view
HBV	Hepatitis B
HCC	Hepatocellular Carcinoma
HCV	Hepatitis C
HVPG	Hepatic Vein Pressure Gradient
INR	International Normalised Ratio
IQR	Interquartile Range
kPa	Kilopascals
LFT	Liver Function Test
LSM	Liver Stiffness Measurement
m/s	Meters / Second
MELD	Model for End-Stage Liver Disease
mm	Millimetres
MMP	Matrix Metalloproteinases
MRI	Magnetic Resonance Imaging
MRE	Magnetic Resonance Elastography
NAFLD	Non-Alcoholic Fatty Liver Disease
NASH	Non-Alcoholic steatohepatitis
NPV	Negative Predictive Value
PACS	Picture Archiving and Communication Service
PBC	Primary Biliary Cholangitis
PPV	Positive Predictive Value
PSC	Primary Sclerosing Cholangitis
q1	Quartile 1

q3	Quartile 3
RTE	Real time Tissue Elastography
ROE	Region of Excitation
ROI	Region of Interest
ROI SD	Standard Deviation of shear wave velocities within the Region of Interest
SD	Standard Deviation
SCD	Skin to Liver Capsule Distance (as per transient elastography literature)
SLD	Skin to Liver Capsule Distance
SR	Success Rate
SSI	Supersonic Shear Imaging
SW	Shear Wave
SWE	Shear Wave Elastography
TE	Transient Elastography
UEC	Urea, Electrolytes and Creatinine
U/S	Ultrasound
VTTQ	Virtual Touch Tissue Quantification
WHO	World Health Organisation

Chapter 1. Introduction

Liver fibrosis is a dynamic process which develops in response to liver injury and repair. It insidiously accumulates over time in patients with chronic liver disease (CLD) and if left unchecked can progress to cirrhosis with its associated complications.[1,2] The degree of liver fibrosis has been shown to correlate with liver disease severity,[3,4] and accurate fibrosis quantification is therefore central to the management of patients with CLD. It has important prognostic implications, helps inform the need and urgency of therapy, and has utility in monitoring treatment response. It is also important for the diagnosis of cirrhosis, which itself necessitates screening for cirrhotic complications.

Liver biopsy remains the gold standard for liver fibrosis assessment, as it has been shown to correlate with future clinical outcome.[3,4] The test can, however, suffer from intra-operator and inter-operator variability and is prone to sampling error due to the small percentage of liver volume analysed.[5–7] This invasive test is painful and also carries a small risk of morbidity and mortality, which further prevent its utilisation in population screening or for temporally monitoring patients over time.[8,9]

A concerted effort has therefore been applied in the development of reliable non-invasive alternatives, to allow physicians to quantify liver fibrosis whilst avoiding the risks of biopsy. Ultrasound based elastography tools are amongst the best-validated and most widely utilised tools in clinical practice. They rely on the assumption that liver stiffness increases in parallel to liver fibrosis severity, and have been shown to be inexpensive and reliable methods of fibrosis quantification. A number of tools have now become available from different manufacturers. The best validated and most widely utilised in the clinical practice include Transient Elastography (Fibroscan®, Echosens, France), Acoustic Radiation Force Impulse elastography (ARFI, Siemens, Germany) and Supersonic Shear Imaging (Aixplorer, France), with a number of new 2D-SWE devices (including Toshiba 2D-SWE) also becoming available over recent years. Whilst these tools have similarities, there are also numerous technological differences which have important implications for clinical practice.

The aim of this Masters was to analyse two different ultrasound shear wave elastography (SWE) systems; Acoustic Radiation Force Impulse imaging (ARFI, Siemens) and 2D-SWE (Toshiba). The two tools are in varying stages of clinical development and use, and the

research questions for the two tools were therefore distinct. The study aims for each system were assessed in two different patient cohorts, and are outlined separately below.

1.1 Acoustic Radiation Force Impulse elastography (ARFI)

ARFI is a point SWE technique which uses a high energy ultrasound push pulse to excite a small volume of tissue within the liver. This results in micrometer tissue displacements away from the area of excitation, whose propagation is monitored using tracking ultrasound beams. The velocity with which these displacements or 'shear waves' move away from the point of excitation is directly proportional to liver stiffness; higher ARFI velocities being indicative of increasing liver fibrosis.

ARFI is now widely utilised in clinical practice and has a number of theoretical and practical advantages over the most extensively evaluated and widely utilised elastography tool, transient elastography (TE). ARFI is performed using a conventional ultrasound machine, which allows the liver to be visualized using B-mode imaging at the time of quantitative assessment. Operators are therefore able to visualize the region of liver being interrogated, ensuring measurements are not inadvertently acquired over vascular or biliary structures. ARFI also uses an automated ultrasound push pulse to achieve tissue excitation, which is theorized to be less-operator dependent than the manual excitation method employed with TE.[10] ARFI is also able to obtain a LSM value in almost all assessed patients, including those with ascites or morbid obesity.[11,12] And most importantly, ARFI has shown equivalent accuracy to Fibroscan® in quantifying liver fibrosis in numerous head-to-head studies and meta-analyses.[13–15]

There is, however, increasing evidence that a number of factors can impact on ARFI performance. Necroinflammatory change,[16–18] right heart failure,[19] post-prandial hepatic congestion,[19,20] and subcapsular measurements taken within 1cm of the liver capsule[21,22] have all been shown to elevate liver stiffness measurements, potentially confounding fibrosis assessment.

Whilst there is increasing validation of ARFI as a clinical tool, a number of questions remain for the technique which have not been fully resolved in the literature.

Firstly, obesity has been linked with reduced ARFI performance in an increasing number of studies.[11,23–26] The magnitude of obesity's impact on ARFI reliability has been variably reported, however, and some studies have failed to demonstrate any significant degradation in performance amongst obese patients.[27–29] In view of these inconsistencies, further evaluation regarding the importance of obesity as a confounding variable is required. In addition to confirming the likely impact of BMI on ARFI performance, we also aimed to evaluate the mechanism underlying this anticipated relationship. Prior studies have postulated that increasing depths of subcutaneous adipose tissue attenuate the ultrasound push pulse; in turn degrading the quality of ARFI LSMs. To test this hypothesis, we aimed to assess whether the thickness of subcutaneous tissue (i.e. skin-to-liver capsule distance, SLD) was more closely associated with ARFI performance than was BMI.

We hypothesized hepatosteatosis would have a similarly negative impact on ARFI reliability. This was based on both anecdotal experience with ARFI, but also on theoretical grounds given the marked ultrasound beam attenuation associated with hepatosteatosis on B-mode imaging.[30] This anticipated relationship has not, however, borne out in a number of small studies to date.[28,29,31,32]

The impact of operator experience on ARFI reliability also remains unknown, being assessed in only two small studies which produced discordant results.[33,34] As a consequence, clinical guidelines do not recommend a minimum operator training requirement for ARFI; which is in contrast to the well established guidelines in place for Fibroscan®.[35]

A further limitation of ARFI is that operators have limited facility to assess the reliability of acquired liver stiffness measurements. Operators are unable to visualize the quality of the shear waves being measured, and clinicians are therefore forced to blindly trust the numerical values generated by the ARFI system. Whilst the degree of spread between the ten measurements obtained within each patient (i.e. IQR/Median ratio) has some utility in gauging the reliability of ARFI results, this approach is imperfect and inaccurate readings can still be seen when IQR/Median ratios are less than 0.30 and vice versa.[31,36] Additional strategies are therefore required to help clinicians better gauge the validity of acquired LSMs. This is of heightened importance given the aforementioned unknowns surrounding the determinants of ARFI performance.

A second research goal was therefore to develop new strategies, which may enable operators to better assess the validity of acquired ARFI measurements. One particular strategy involved routinely scanning patients with multiple independent operators and using the degree of operator agreement as a marker of ARFI reliability. We hypothesized that the presence of inter-operator discordance may herald unreliable ARFI results.

Finally, whilst ARFI has demonstrated consistently high accuracy in liver fibrosis quantification, more variable performance has been reported in the assessment of cirrhosis severity. Some groups have shown very high accuracy in this setting, particularly in identifying patients with portal hypertension[37,38] and oesophageal varices,[39–41] whilst others have reported only modest associations.[42,43,44,45] Hence further clarification is again required.

Hypotheses

1. Obesity is anticipated to be a key determinant of ARFI performance, which is hypothesized to be secondary to increased attenuation from subcutaneous adipose tissue. ARFI reliability is therefore anticipated to have a stronger association with SLD than BMI.
2. Hepatosteatosis is also hypothesized to have a strong negative impact on ARFI reliability.
3. ARFI performance is theorised to be relatively operator independent. We therefore anticipate the technology will show minimal training effect, both at an individual operator and institutional level.
4. ARFI LSM is hypothesized to have a strong association with the severity of cirrhosis. In particular, ARFI is anticipated to show high performance in predicting the presence of cirrhotic complications and to show strong associations with prognostic indices.
5. Inter-operator agreement is hypothesized to represent a useful indicator of ARFI reliability. We anticipate ARFI will show poor accuracy amongst patients in whom operators obtain discordant LSM values.

Aims

1. Assess the strength of relationship between obesity and ARFI performance measures including IQR/Median, inter-operator agreement and accuracy. Furthermore, to determine whether SLD or BMI is more closely associated ARFI reliability.
2. Determine whether hepatosteatosis has an association with the ARFI performance measures mentioned above.
3. Assess whether ARFI performance changes over time with increasing operator experience. This includes with increasing number of scans performed per operator, but also for the institution overall.
4. To assess the performance of ARFI in assessing cirrhosis severity. In particular, its association with prognostic indices (i.e. Child Pugh and MELD scores) and accuracy in predicting the presence of cirrhotic complications (i.e. encephalopathy, ascites and oesophageal varices).
5. To determine whether the presence of inter-operator discordance is associated with reduced ARFI accuracy. And if so, to identify which group of patients would receive greatest benefit from the new reliability assessment strategy.

1.2 Toshiba 2D-SWE

2D-SWE is the most recent and technologically advanced addition to the ultrasound elastography armamentarium. It is similar to the Siemens ARFI device in the use of acoustic radiation force 'push pulses' to generate shear waves in the liver. 2D-SWE however uses multiple near simultaneous pulses to interrogate a larger region of liver, rather than just a small 'point' of tissue encompassed within the measurement region of interest (ROI). This allows 2D-SWE to construct a two-dimensional map of shear wave propagation throughout a section of liver, enabling operators to visualize and qualitatively assess regional shear wave characteristics. A quantitative elasticity measurement can then be obtained by positioning a measurement ROI in an area deemed suitable for assessment.

2D-SWE has a number of theoretical advantages over related ultrasound-based elastography techniques, including ARFI. The ability to visualize shear wave propagation

firstly allows operators to qualitatively assess the reliability of single shot acquisitions. This enables operators to not only reject acquisitions which are prone to artefact, but also to position the measurement ROI in areas with optimal shear wave characteristics. Supersonic Shear Imaging (SSI) is the best validated 2D-SWE device in clinical use, and early data has found SSI to be equivalent or superior to TE and ARFI in the quantification of liver fibrosis.[24,46,47] SSI has also shown very high internal consistency between measurements, and as a result current clinical guidelines suggest as few as three readings may be required to provide a precise estimate of liver stiffness.[35]

A new 2D-SWE technology has been developed by Toshiba Medical Corporation (Tochigi, Japan). This device is similar to the SSI system, but utilises novel processing methods to generate a second display mode termed the 'Propagation Map'. This illustrates shear wave arrival times at different points in the liver as contour lines; information which is different and purportedly complementary to the 'Speed Smart Map' display, which is ubiquitous amongst other 2D-SWE systems. This additional display mode is hypothesized to allow better evaluation of regional shear wave propagation characteristics above that of the Speed Smart Map, which may in turn aid in the optimisation of ROI positioning and thereby impact on 2D-SWE measurement reliability. The device is in the very early clinical phases of testing and there is currently minimal published data for the Toshiba system. As a consequence, no measurement acquisition guidelines have been established for this technique to date. Whilst data for related elastography tools provide a starting point, there are several unknowns which prevent their direct implementation for the Toshiba 2D-SWE system. The aim of the second half of the masters was therefore to evaluate a number of fundamental technical parameters, to assist in the formation of Toshiba 2D-SWE acquisitions guidelines. The development of such guidelines is important not only for the optimisation of 2D-SWE performance in clinical practice, but also for the standardisation of acquisition protocols in future clinical trials.

Our primary research focus was on the internal consistency of 2D-SWE measurements. We hypothesized that Toshiba 2D-SWE's novel Propagation Map would allow operators to better select a region of liver parenchyma suitable for quantitative assessment, thereby yielding more consistent and more reliable 2D-SWE measurements. We anticipated that a low variability in measurements (i.e. low IQR/Median) would therefore be observed for each patient, and that fewer measurements would therefore be required per patient to provide a precise estimate of liver stiffness.

We also aimed to assess factors affecting 2D-SWE measurement consistency, which may forewarn of potential issues with 2D-SWE reliability and accuracy. Given the shared technology underpinning the ARFI and Toshiba 2D-SWE platforms, it was hypothesized that factors affecting ARFI would have a similar impact on the new technology. Areas of interest included obesity, SLD, hepatosteatosis and measurement depth within the liver.

The final research focus was to evaluate ROI SD/Speed as a possible indicator of individual 2D-SWE measurement reliability. 2D-SWE is unique in providing a standard deviation of shear wave velocities recorded within the measurement ROI (i.e. ROI SD) for each individual LSM. Measurement ROIs with a heterogeneous shear wave profile are reflected by higher ROI SD values and vice versa. Given that the variability of LSMs obtained within each measurement set (i.e. IQR/Median) has been shown to reflect the set's overall reliability, we hypothesized that similar principles may have relevance at the individual measurement level. We hypothesized that measurements with relatively uniform shear wave velocities (i.e. low ROI SD/Speed) would yield more reliable results than those with more heterogeneous profiles. Our aim was therefore to evaluate ROI SD/Speed as a potential indicator of individual measurement reliability.

Hypotheses

1. Toshiba 2D-SWE is hypothesized to have high internal measurement consistency (i.e. low IQR/Median ratios) relative to other widely utilised elastography tools, including ARFI.
2. Fewer 2D-SWE measurements may be required per patient to provide a precise estimate of liver stiffness.
3. Factors affecting ARFI reliability are hypothesized to have a similar effect on 2D-SWE. We therefore anticipate increased body habitus (i.e. BMI / SLD) and hepatosteatosis will be associated with higher IQR/Median ratios.
4. The ROI SD/Speed may represent a useful indicator of individual measurement reliability. Specifically, measurements with higher ROI SD/Speed values are hypothesized to show poorer measurement consistency.

Primary aims

1. Assess the measurement variability (i.e. IQR/Median) of Toshiba 2D-SWE.
2. Determine the association between patient factors (particularly BMI, SLD and hepatosteatosis) and IQR/Median.
3. Determine the number of 2D-SWE measurements required per patient to provide a reliable liver stiffness estimate (i.e. within 5% of the existing method using ten samples).
4. Assess the relationship between the ROI SD/Speed and the internal consistency of individual 2D-SWE measurements.

Chapter 2 Background

2.1 Liver fibrosis

2.1.1 Pathogenesis

Liver fibrosis or 'scarring' is a dynamic process which occurs as a healing response to liver injury.[48] It can be triggered by a wide spectrum of toxic insults ranging from viral pathogens, metabolic abnormalities (including hepatosteatosis), autoimmune diseases or exogenous toxins (e.g. alcohol, pharmaceutical agents). Whilst the potential stimuli are diverse, they trigger a common pathologic process which culminate in the deposition of extra-cellular matrix (ECM) to encapsulate and isolate the insult from the remaining liver parenchyma. Whilst the production and degradation of ECM are kept in close equilibrium in the healthy liver, the persistent activation of fibrogenesis in chronic liver disease (CLD) can distort this balance and lead to the gradual accumulation of fibrosis.

The pathologic process underlying liver fibrosis is complicated and involves numerous chemical mediators and cell types.[49] The central player is the stellate cell or myofibroblast, which is the primary cell type responsible for collagen formation.[50] These ordinarily reside in a quiescent, retinoid containing state, however become activated in the setting of acute or chronic liver injury through the release of a wide range of pro-inflammatory and pro-fibrogenic mediators including oxygen free radicals, TNF-alpha, IL-1 and platelet derived growth factor (PDGF).[51] Once activated, myofibroblasts proliferate and start depositing increased quantities of ECM composed of collagen, glycoproteins, glycosaminoglycans and proteoglycans. The ECM composition also shifts with inflammation / activation, with the myofibroblasts depositing fibril-forming collagen subtypes (i.e. types I and III) in preference to type IV.[52] These collagen subtypes have a propensity to become highly cross-linked, which makes them resistant to degradation by matrix metalloproteinases (MMPs) and further contributes to the disequilibrium in fibrosis formation / degradation.

Liver inflammation, stellate cell activation and fibrosis accumulation have a number of pathologic consequences. The changes firstly damage and disrupt the function of surrounding hepatocytes and endothelial cells; thereby impacting on liver function. The abnormal ECM deposition results in sinusoid remodeling, septa formation and distortion of the overall liver architecture, which can ultimately lead to formation of regenerative liver

nodules, the hallmark of cirrhosis. Myofibroblasts also have smooth muscle like properties and contain internal contractile filaments. On activation, these can distort the liver architecture and increase portal resistance, further contributing to portal hypertension. Finally liver inflammation / fibrosis represents a pre-cancerous state, as it provides a microenvironment which facilitates the development of primary liver cancers, especially hepatocellular carcinoma (HCC).

2.1.2 Natural history

The accumulation of liver fibrosis is a chronic process which occurs over months to years in response to persistent insult. The process occurs insidiously without symptoms, and is not ordinarily associated with complications in its early stages due the liver's significant physiologic reserve.[53]

Historically, the development of liver fibrosis was believed to be a unidirectional and irreversible process, with the best hope being to halt its accumulation.[54] This dogma has however been proven false, and it has been well demonstrated that liver fibrosis can regress following the removal of the toxic insult.[55] The potential reversibility of lower levels of fibrosis (i.e. Metavir Scores F0 – F3) has been demonstrated for almost all CLD aetiologies including following the institution of anti-viral therapy in both Hepatitis B,[56] and Hepatitis C,[57,58] immunosuppression in autoimmune hepatitis (AIH),[59] weight loss in NAFLD,[60] and venesection in hereditary haemochromatosis.[61] There is now also increasing recognition that patients with established cirrhosis may even regress during its earlier stages.[62,63] The exact point when fibrosis becomes irreversible is however unknown, and is likely to be dependent on a number of undefined patient and disease factors.

The slow pace and potential reversibility of liver fibrosis therefore provides a large window and great opportunity for therapeutic intervention. If left unchecked however, CLD can gradually progress to cirrhosis with its associated complications including synthetic failure, portal hypertension and hepatocellular carcinoma (discussed further below).[2,3]

2.1.3 Clinical importance

The extent of liver fibrosis has been shown to be proportional to the severity of CLD and underlying hepatic reserve. The quantification of liver fibrosis therefore provides valuable prognostic information regarding the clinical status of a patient with CLD, with the degree of fibrosis being shown to predict the risk of synthetic failure, development of portal hypertensive complications and ultimately death.[3,4] The prognostic implications of liver fibrosis are widely relied upon for the clinical management of CLD patients.[64] This includes but is not limited to screening patients at high-risk of developing CLD and liver fibrosis (e.g. those taking hepatotoxic medications), triaging the need or urgency to initiate treatment (particularly in the case of viral hepatitis), assessing the clinical progress of patients both prior to and following therapy commencement, as well as diagnosing cirrhosis.

Identifying cirrhosis is considered the most important fibrosis endpoint in clinical practice guidelines.[65] Cirrhosis can be categorized as either compensated or decompensated, and can be associated with a wide range of complications including portal hypertension, ascites, encephalopathy, gastric and oesophageal varices and synthetic failure (coagulopathy, hypoalbuminaemia, etc.). The diagnosis therefore confers a negative prognosis, with a 2 year survival rate of 74% and 34% amongst patients with compensated and decompensated cirrhosis respectively.[66] The diagnosis of cirrhosis not only has implications regarding the management of their underlying CLD, but also mandates endoscopic surveillance for oesophageal varices and screening for HCC using ultrasound ± alpha-fetoprotein levels in the appropriate clinical context.

Whilst cirrhosis remains widely used in the hepatology vernacular, the ongoing use of the term has been recently challenged due to increasing recognition of its deficiencies as a concept.[67] Firstly, cirrhosis is an over simplistic term which indiscriminately groups patients with wide ranging CLD aetiologies, disease severities and prognoses under a single banner. The morphological changes of cirrhosis are not of singular importance, with HVPG (hepatic venous pressure gradient) and continuous measures of liver fibrosis severity (including elastography) also having relevance to patient care. And finally the term has negative connotations in the wider community, being viewed as an end-stage and irreversible clinical state which is becoming increasingly inaccurate with advent of improved therapies. As a consequence, there have been recommendations to simply

describe a particular fibrosis stage for a patient and a push to replace 'cirrhosis' with the term 'advanced fibrosis'.

The importance of accurately quantifying liver fibrosis is further heightened by the large scale of CLD as a health issue; cirrhosis currently being among the 10 most common causes of death worldwide.[68] It is also equally relevant to both developed and developing countries due to epidemic levels of viral hepatitis and NAFLD, respectively. And the health burden is expected to only worsen, as the prevalence of NALFD continues to rise worldwide.[69] NALFD is estimated to affect as many as 25 – 35% of the population in the United States, and has been identified as one of the major future health challenges.[70]

2.2 Liver Biopsy

Liver biopsy has long been regarded as the gold standard in fibrosis assessment, and remains the accepted reference in both clinical practice and research. This is largely due to the extensive long-term data showing that the severity of fibrosis on biopsy (referred to as F score) closely predicts patient outcome and the future development of CLD complications.[3,4] Despite this, liver biopsy has a number of limitations which make it an imperfect gold standard.

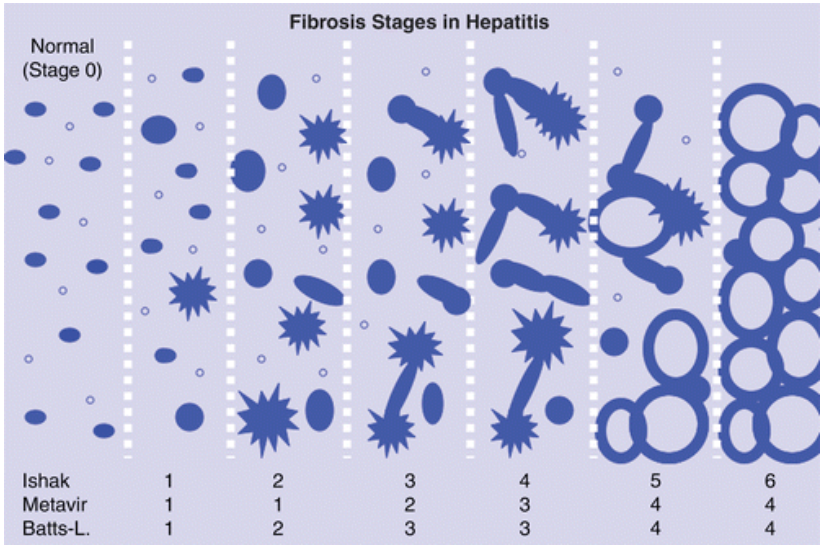
2.2.1 The Imperfect gold standard

The technique is firstly prone to sampling error, as only a very small portion of liver (approximately 1/50,000th) is ordinarily obtained for assessment.[5–7] This issue is further compounded by the often patchy fibrosis observed in chronic liver disease, particularly cirrhosis. As a consequence, it is imperative that biopsy samples are of good quality and adequate size, with longer specimens shown to improve the accuracy of assessment. It is therefore recommend that a sample be at least 15mm in length and contain at least 6 portal tracts to help minimise sampling error.[71,72]

The interpretation of liver fibrosis is also highly subjective and therefore prone to significant intra and inter-operator variability.[73] Overall agreement between pathologists has shown to be only moderate, with one study observing at least one F score difference between pathologists in 25% of patients.[74]

The categorical grading systems used are also imperfect, as they do not truly describe the continuous scale of liver fibrosis severity observed across the spectrum of chronic liver disease. This is particularly the case for cirrhosis, with many histopathology scoring systems failing to differentiate between early and advanced cirrhosis. The grading systems are also not well suited to temporal assessment, as the small number of arbitrary fibrosis categories limit the sensitivity and accuracy of tracking patient progress over time. The wide variation in scoring systems used internationally also provides further complexity to the standardisation of biopsy results in both clinical research and practice (Figure 2.1 below).

Figure 2.1. Common liver fibrosis grading systems, demonstrating the variability in F score meaning across the spectrum of liver fibrosis severity. Image from Goodman et al.[75]



2.2.2 Risks and complications

The other primary limitation of liver biopsy is its invasive nature and associated risks of morbidity and mortality. The procedural mortality rate is widely quoted to be approximately 0.01% and 3% of patients require hospitalization for complications arising from the procedure.[8,9,76] The most common side effect is pain, with significant episodes reported in 25% of patients.[77] More serious adverse events also include serious bleeding (0.3% of patients), hypotension, bile peritonitis, bacteraemia and visceral perforation.

Whilst many of these side effects are either transient or rare, the risks are nonetheless real and need to be justified and carefully considered before a liver biopsy is performed.

2.2.3 Current role of liver biopsy

The numerous limitations of liver biopsy have seen its use decline over recent years. A number of traditional indications of biopsy have now been replaced by non-invasive methods, which continue to increase in availability and performance. This is particularly the case for screening high-risk populations (e.g. patients with NAFLD or taking hepatotoxic drugs) and for the temporal monitoring of patients; in which the risks of biopsy may not be justifiable or acceptable to patients.

Liver biopsy nonetheless provides a wealth of ancillary information regarding liver status, which is not provided by other assessment methods. This includes the presence of inflammation, hepatosteatosis, cholestasis, mineral deposition and vascular congestion. These factors not only allow for the assessment of underlying disease aetiology, but also help to account for confounders which may potentially impact on the reliability of non-invasive assessment of fibrosis. Therefore whilst the use of liver biopsy declines, the technique will likely retain an important role in both clinical and research settings for the foreseeable future.[78]

2.2.4 Hepatic venous pressure gradient (HVPG)

HVPG is another invasive test which can help inform the severity of advanced fibrosis and cirrhosis. It involves measuring the pressure difference between the hepatic and portal veins using a venous catheter; the latter approximated by the wedged hepatic venous pressure. The trans-hepatic pressure gradient increases with hepatic fibrosis and the disruption of sinusoidal liver architecture and is used to define the presence of portal hypertension (HVPG >5mmHg) and clinically significant portal hypertension (HVPG >10mmHg).[79] The test has shown high utility in predicting the risk of developing complications (especially oesophageal varices), HCC and overall survival.[79]

2.3 Non-invasive fibrosis assessment

The non-invasive assessment of liver fibrosis is currently achieved through two primary methods; namely serum markers and elastography tools. These tools do not directly assess liver fibrosis, and instead are surrogate markers measuring the secondary byproducts or physiologic changes associated with fibrosis development. As a consequence, both methods have their relative strengths and weaknesses, with their accuracy being dependent on a number of factors which are discussed below.

2.3.1 Serum markers

Serum tools are in widespread clinical use for fibrosis assessment, and continue to increase in number as our knowledge surrounding the pathogenesis of liver fibrosis expands.[80] The tests can be broadly categorized as either direct (Class I) or indirect (Class II) biomarkers. Direct markers measure components directly involved in ECM turnover, which often incorporate specialised assays available in limited laboratories. They are considered the most sophisticated and accurate of the serologic tools, with the most validated examples including FibroTest,[81] Hepascore,[82] FibroSpect,[83] and the European Liver Fibrosis Study Group panel.[84] Indirect markers assess the secondary effects of fibrosis, including hepatic function, portal hypertension or necroinflammatory change. The tools are mostly built upon commonly available assays, and are therefore more widely available / inexpensive but of generally lower accuracy than their counterparts. The most validated examples of indirect tools include the APRI index,[85] FIB-4 score,[86] and AST/ALT ratio.[87]

Serum tools have the benefit of being highly reproducible, quantifiable, objective and are easy to perform; which makes them well suited to the serial monitoring of patients. They do however have a number of inherent limitations which prevent their standalone assessment of liver fibrosis.[88] Serum markers firstly evaluate matrix turnover rather than severity, and can therefore be falsely positive in cases of active hepatitis or conversely miss significant fibrosis in the absence of inflammation. Some serum markers are not liver specific, and can be elevated by extra-hepatic sites of inflammation. Serum levels also depend on clearance, and can therefore also be impacted by hepatic and renal function.

These issues contribute to serum markers having only modest accuracy in the assessment of liver fibrosis; with a meta-analysis of the most common serum panels showing a median

accuracy of AUROC = 0.82 in assessing HCV.[89] The tools have shown particularly poor accuracy in differentiating between moderate levels of fibrosis, and a large majority of patients receive indeterminate results.[90] Furthermore no serum panel has emerged as standard of care, resulting in variable application of different tools according to local preference and availability.

2.4 Elastography

Elastography involves the evaluation of tissue stiffness as a marker of disease. In more technical terms, the European Federation of Societies of Ultrasound in Medicine and Biology (EFSUMB) describe elastography as the assessment of the "biochemical properties associated with the elastic restoring forces in the tissue that act against shear deformation".[35] Elastography tools all rely on a common underlying approach, which involves assessing the degree of tissue deformation generated by an applied excitatory force. The tissue deformation characteristics can then be used to define the tissue's Young's modulus, which is the physical parameter used to define stiffness.

In liver fibrosis assessment, elastography relies on the principle that liver stiffness increases in proportion to fibrosis severity.[91,92] Whilst this premise holds true, there is increasing recognition of numerous additional factors which can also impact on liver stiffness. The liver is surrounded by a minimally distensible capsule, and therefore any factor which increases liver volume will similarly increase tissue stiffness.[93] These factors are discussed in further detail in below sections, however include the post-prandial state,[19,20] deep breath holds (with secondary valsalva effects),[94,95] hepatic inflammation,[16–18] right heart failure,[19] and cholestasis.[96] These factors interfere with the premise underpinning the use of elastography in liver fibrosis assessment, and are therefore a shared limitation of all elastography techniques.

There are now numerous elastography technologies in clinical use, which employ a broad range of approaches. These technologies can be sub-classified in a number of ways, including by the underlying physical principle being assessed (i.e. strain vs. shear wave), the method of excitation employed, the method of tissue tracking used, as well as differences in processing and/or display. These are briefly outlined below.

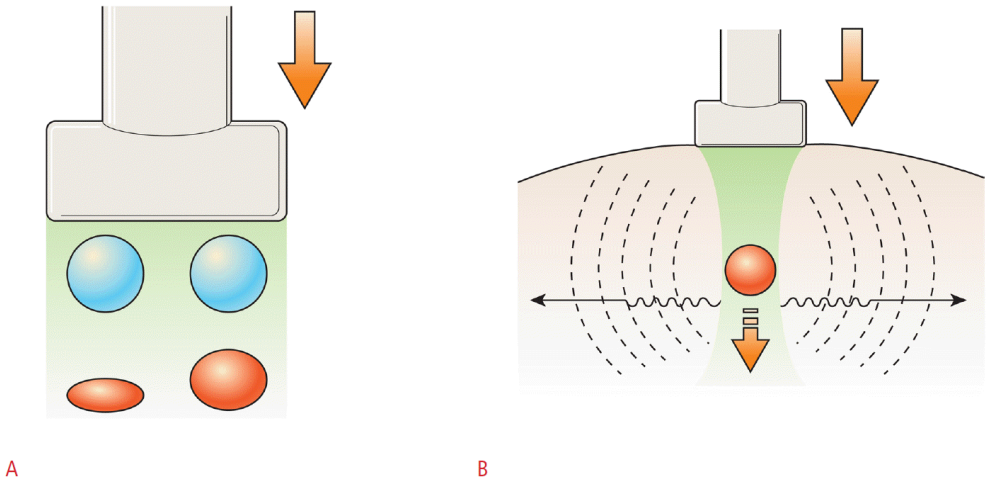
2.4.1 Strain vs. Shear Wave Elastography

Strain Elastography

Strain is the oldest form of elastography and assesses tissue compressibility, most commonly through the use of repeated axial compressions via an ultrasound probe (Figure 2.2). Sequential ultrasound images are then used to assess relative tissue deformation, which is then converted into a qualitative elastogram or 'strain image'. The most widely used strain devices include Real-time Tissue Elastography (RTE, Hitachi) and eSie Touch Elastography Imaging (Siemens).

Tissue strain rapidly decays with depth and is therefore primarily suited to the assessment of superficial lesions (e.g. within the thyroid or breast). The liver is difficult to assess, as the chest wall limits the direct application of manual probe compression. Strain imaging also assesses relative rather than absolute tissue compressibility, and is therefore predominantly qualitative rather than quantitative. For these reasons, strain elastography is not currently recommended for the use of liver fibrosis assessment,[35] and is therefore not discussed further.

Figure 2.2: Illustration of Strain Elastography (A) and Shear Wave Elastography (B). Strain looks at the compressibility of tissue in the axial plane relative to surrounding structures; displacement being greater with soft tissue than hard tissues. SWE measures tissue displacement (i.e. shear waves) which propagate perpendicularly away from the area of excitation. The speed of shear wave propagation can then be used to estimate tissue stiffness. Image from Kwak et al.[97]



A B

Shear Wave Elastography (SWE)

Shear Wave Elastography (SWE) encompasses the majority of technologies used clinically for liver fibrosis assessment, including transient elastography (i.e. Fibroscan®), point SWE, two-dimensional SWE (2D-SWE) and magnetic resonance elastography (MRE). Shear is defined as the change in shape of an object following excitation, which occurs without a change in the object's volume. This shape change results in the displacement of surrounding tissues, which propagate away from the area of excitation as shear waves. SWE measures the arrival of these shear waves at locations around the point of excitation, the characteristics of which are dependent on the tissue's elasticity properties. In stiffer materials, shear waves propagate at high speed but are also very quickly attenuated and therefore travel only short distances. In contrast, shear waves propagate at a slower velocity in softer materials, recover more slowly and can potentially travel over long distances. Hence the speed of shear wave propagation and recovery are proportional to tissue stiffness, and can thereby be used as a biomarker of liver fibrosis.[98]

2.4.2 Excitation Method

Elastography assesses tissue response to excitatory forces, and the method used to achieve excitation differs between the technologies. The excitation techniques employed can be broadly categorized as either 'quasi-static' or 'dynamic'.

Quasi-static methods

Quasi-static methods involve the application of a constant or slowly changing stress to achieve tissue deformation; the most prominent examples including probe pressure and physiologic movement (i.e. from heart contractility, lung movement, arterial pulsatility or the movement of skeletal muscles). Similar to strain imaging, quasi-static excitation methods are not well suited to liver fibrosis assessment. The chest wall again inhibits the most common method of quasistatic excitation (i.e. probe pressure), and quasi-static excitation relies on the assessment of the relative compressibility of tissue and is therefore primarily qualitative. Whilst some quantitative analyses can be achieved through comparison with a reference tissue / phantom, the level of quantitation is insufficient for clinical purposes. Therefore dynamic excitation is the primary method employed in the quantitation of the liver fibrosis.

Dynamic methods

Dynamic excitation involves the application of a time varying force to a tissue, which results in rapid changes in tissue deformation. This can be achieved through the use of vibrations (e.g. TE and MRE) or impulses (point SWE and 2D-SWE) as described below, however both approaches result in the generation of shear waves. The greater amplitude in tissue deformation achieved with dynamic excitation allows more accurate quantification of tissue elasticity than quasi-static methods.

Mechanical excitations methods primarily use vibrations to generate shear waves in a tissue. Notable examples are TE and MRE, which both achieve manual excitation through use of an external vibratory probe. These methods require shear waves to propagate from the skin surface deep into the liver, and are therefore designed to generate very low velocity shear waves (within range of tens of Hz) to help minimize tissue attenuation. The second dynamic technique, Acoustic Radiation Force Impulse (ARFI), is the excitation method underpinning ARFI imaging, point SWE and 2D-SWE.[99] ARFI utilises a high-energy ultrasound 'push pulse' to excite a small, localised area of tissue remote from the ultrasound probe. It does this by focusing low frequency but high intensity ultrasound waves at a target location, whose longer wavelengths (i.e. 50-1000 us) are specially chosen for their high energy transfer / absorption characteristics. The imparted force is proportional to the intensity of the ultrasound pulse and the absorption co-efficient of the tissue, and results in micron level displacements of the tissue according to Hooke's Law.[100]

ARFI excitation can be utilised in two different ways to assess tissue elasticity (Figure 2.3). The original approach developed at Duke University measured the axial displacement caused by ARFI, which occurs in the direction parallel to the push pulse. These 'compression waves' can be tracked to produce a qualitative image of relative tissue compressibility / stiffness. This is the technologic principle used in ARFI imaging.

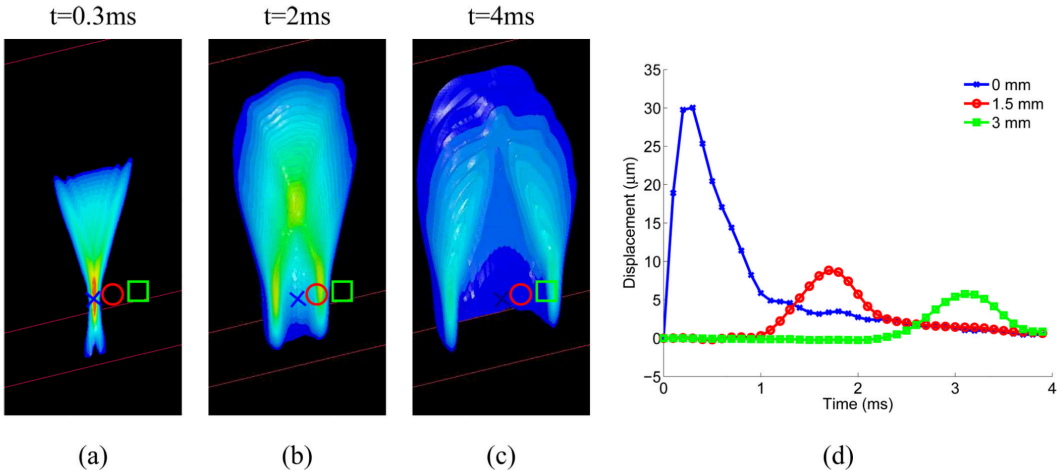
The second approach relies on the finding that some of the axial compression waves are converted into shear waves, which are micron level displacements that propagate in a perpendicular direction away from the region of excitation (ROE). Significant processing is required to differentiate shear waves from the overlapping compression waves, nonetheless the speed of shear wave propagation can be measured and quantified to provide an estimate of the tissue shear wave modulus.[101] The use of ARFI to generate

shear waves is relied upon by the two elastography technologies evaluated in this thesis; namely ARFI and Toshiba 2D-SWE.

ARFI has a number of theoretical benefits over alternative excitation techniques. Firstly it allows focused evaluation of a small area of selected of tissue, including in deep locations remote from the transducer. It can excite tissues which cannot be manually compressed due to overlying structures (e.g. the costal wall), which make it particularly well suited to liver assessment. And finally, the technique involves a standardized excitation impulse which is automatically generated by the ultrasound device. The technique therefore uses less manual processing, and on a theoretical level should be less-operator dependent than other excitation methods.[10]

There are however a number of technical variables beyond underlying tissue stiffness which can impact on the velocity of shear waves generated by the ARFI technique. Firstly the frequency of the push pulse can impact on shear wave generation, with lower frequency pulses paradoxically resulting in greater energy transfer and thereby higher shear wave velocities.[21] Evaluating deeper tissues results in increased attenuation of the push pulse, lower energy transfer and therefore lower shear wave velocities.[21,98,102] The angle of the ROI is also believed to be important, with oblique angles being associated with increased beam refraction and thereby lower ARFI excitation.[103] Finally, anatomic factors can also impact on shear wave generation and propagation, either through the presence of internal macroscopic structures or anisotropy (i.e. direction dependence) of the tissue. These latter anatomic factors are however less relevant to the liver, given its relative homogeneity and isotropic structure, and are most pertinent to the evaluation of complex organs such as the kidneys.[103]

Figure 2.3. Image of an ARFI impulse, demonstrating tissue displacement at three different time points following excitation (i.e. 0.3, 2 and 4ms). The images show displacement in the axial direction (i.e. towards the top of the figure), which is measured in ARFI imaging. For quantitative analysis, the speed of shear wave propagation in a perpendicular direction is measured (i.e. towards the red circle and green square). Image from Nightingale et al.[104]



2.4.3 Imaging method

Following excitation, tissue deformation can be monitored through a number of techniques including ultrasound, MRI sequences (i.e. MRE) and less commonly pressure / stress sensors in Tactile Imaging. Ultrasound is particularly well suited to this purpose, and is therefore the most widely employed tracking method in clinical practice. Ultrasound waves firstly travel at much higher speeds than shear waves, which allows them to track and measure shear wave propagation very accurately. There are also techniques to track the speckle noise in an ultrasound image, which allows ultrasound to track tissue displacement even in the absence of identifiable sonographic structures.[105] Ultrasound furthermore has the capability for real time imaging, is less expensive and also more widely available than magnetic resonance techniques.

2.4.4 Processing / Display

A further area of difference between available elastography tools relates to their processing methods and display options. From a practical perspective, the greatest difference is the

variable provision of qualitative vs. quantitative information. Some elastography tools such as strain imaging and ARFI imaging are purely qualitative, and provide an image of tissue elasticity (also known as an 'elastogram') but without any quantitative information. Other tools including point SWE (i.e. ARFI elastography) and TE are essentially quantitative, and provide an elasticity measurement without a clinically useful elastogram. MRE and some 2D-SWE devices however provide both.

Another ostensible difference between devices is the units used to describe tissue elasticity. The true unit of tissue stiffness is the Young's modulus, which is measured in kilopascals (kPa). In practice, however, a range of measurement units are used by the different techniques, which limits comparison and cross-referencing between technologies. Strain imaging, TE and SSI report their results in kPa, MRE use the shear modulus, whilst point SWE and many other 2D-SWE techniques report shear wave propagation in meters per second (m/s). Point SWE and 2D-SWE systems can all convert their shear wave velocities into kPa using mathematical formulae, as is the case for TE and SSI. This conversion formula however makes assumptions regarding tissue density (assuming a density of 1g/mL) and homogeneity. Many recent systems therefore report their elasticity measurements as a velocity (m/s) rather than kPa, to help improve standardisation and comparison between techniques.

2.5 Elastography tools for liver fibrosis assessment

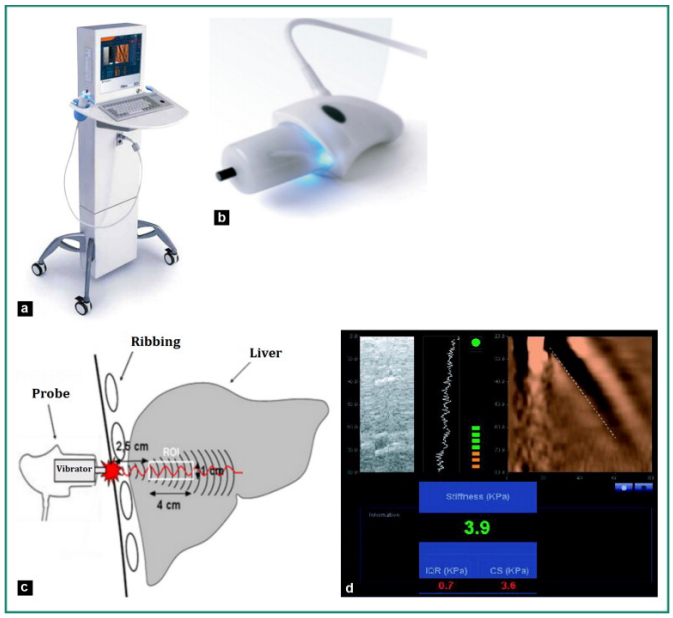
Amongst the wide spectrum of available elastography techniques, there are a number of devices which have been validated and are now in wide clinical use for the evaluation of diffuse liver disease. The most established techniques including Transient Elastography (TE), Magnetic Resonance Elastography (MRE), point SWE and 2D-SWE are discussed below.

2.5.1 Transient Elastography

Whilst all SWE techniques involve a form of transient elastography (TE), this term has become synonymous with the Fibroscan® device developed by Echosens (Paris, France).[106,107] TE remains the most widely utilised and validated elastography tool in liver

fibrosis assessment, and therefore is an appropriate historical reference for other elastography technologies.

Figure 2.4: Images of the Transient Elastography technique; including the ultrasound machine (a), probe with inbuilt mechanical piston (b), diagram of shear wave generation / propagation into the liver (c) and a TE screenshot (d). Image from Frulio et al.[108]



In basic terms, the device involves an ultrasound transducer built onto the axis of a mechanical vibrator. Pushing a button causes an automated piston to generate low-frequency vibrations to propagate into the underlying tissues / liver. A pulse-echo ultrasound beam is then used to measure the velocity of the generated shear waves, which are mathematically converted into a Young's Modulus (kPa). There are two probes available for the device; the standard M probe (3.5 MHz), and a lower energy XL probe (2.5 MHz) which is designed for use in obese individuals.

There is now extensive literature supporting the use of TE in a wide range of chronic liver diseases, with numerous meta-analyses reporting high accuracy in the assessment of HCV,[109,110] HBV,[111,112] and NAFLD.[113] A meta-analysis incorporating 50 studies found a mean AUROC of 0.84 (95%CI: 0.82 – 0.86), 0.89 (95%CI: 0.88 – 0.91) and 0.94 (95%CI: 0.93 – 0.95) in diagnosing significant fibrosis (≥F2), severe fibrosis (≥F3) and cirrhosis (F4), respectively.[114] These findings are reflected in elastography guidelines, which have stated

that Fibroscan® has greatest accuracy in the diagnosis of cirrhosis and severe fibrosis, but has more limited utility in assessing the presence of lower fibrosis levels.[93]

Transient Elastography is also relatively unique in the strong longitudinal data available from its years of clinical use. TE has shown power in predicting future risk of hepatic decompensation,[115] liver cancer,[116] and long-term patient survival[117,118]; with one study even suggesting TE to be superior to liver biopsy in predicting 5 year survival in patients with HCV.[119] The device has furthermore shown utility in temporally tracking the response to therapy in both HCV[120–122] and HBV.[123–125]

The use of the IQR/Median ratio as a predictor of scan reliability was also founded and refined for TE, before being widely adopted by analogous elastography systems. The reliability of TE was originally shown to be reduced amongst scans with a success rate (SR) <60% or an IQR/Median >30%, and these cut-offs were subsequently applied as indicators of poor TE reliability.[106,126,127] These criteria have since been refined, with SR being omitted from the criteria after being found to be of lesser importance than the IQR/Median ratio. A further adaption has been the inclusion of a minimum LSM cut-off threshold of >7.1 kPa.[128] This acknowledges that TE is predominantly associated with false positive results, and that readings with low LSMs are therefore likely to be accurate irrespective of the IQR/Median ratio. And most recently, Boursier *et al.* has suggested that IQR/Median may be further sub-categorised into <0.10, 0.10 – 0.30 and >0.30 brackets to further improve the stratification of TE reliability.[129]

Transient elastography does however have a number of well recognised limitations. Firstly, the device is unable to obtain a valid reading in up to 20% of patients, particularly in the setting of obesity and the metabolic syndrome.[12,130] Obesity has also been associated with reduced TE accuracy; which has only partially been overcome through use of the XL probe.[12,23,131] TE reliability is affected by ascites, which is thought to impede shear wave propagation.[107] And performance is also affected by a wide range of factors impacting on underlying liver stiffness; including cholestasis,[96] right heart failure with subsequent hepatic congestion,[132] and hepatic inflammation.[133,134]

On a more practical level, TE requires the purchase of a dedicated machine which cannot perform traditional ultrasound examination. It lacks an associated B mode image, and measurements are therefore acquired blindly without knowledge of the underlying liver

structures. And finally it is highly operator dependent, with a long period of operator training (i.e. 500 scans) required to be considered an expert in the technique.[12,35]

2.5.2 Magnetic Resonance Elastography (MRE)

Magnetic resonance elastography (MRE) is another well-validated tool. It too relies on an external mechanical vibrator to generate shear waves, however measures propagation characteristics using the MRI spin echo sequence.[135]

MRE has demonstrated very high accuracy in a number of meta-analyses,[136,137] with some head-to-head studies reporting MRE to be the most accurate of the elastography tools in current clinical practice.[138,139] MRE also has the advantage of very high acquisition success rates, relatively uniform performance across operators, and also allows 3D assessment of the entire liver; rather than being restricted to areas accessible via an acoustic window.

Despite the excellent accuracy of MRE, the technology's expense and limited availability have hindered its broad application in clinical practice. MRE has limited accuracy in iron overload states due to interference with low signal-to-noise. Also the small gantry can pose issues for patients with claustrophobia or morbid obesity; the latter becoming of increasing relevance in the context of the growing NAFLD epidemic.

2.6 Point shear wave elastography (point SWE)

Shear Wave Elastography utilizing ARFI excitation methods are broadly categorised as point SWE, where a very small volume or 'point' of tissue is studied, and two-dimensional (2D) SWE where the elasticity profile of a larger section of tissue is evaluated.[140] The very localized, micron level displacements generated in point SWE are insufficient to reconstruct an elastogram (i.e. elasticity image), and the technique provides purely quantitative information regarding tissue stiffness.

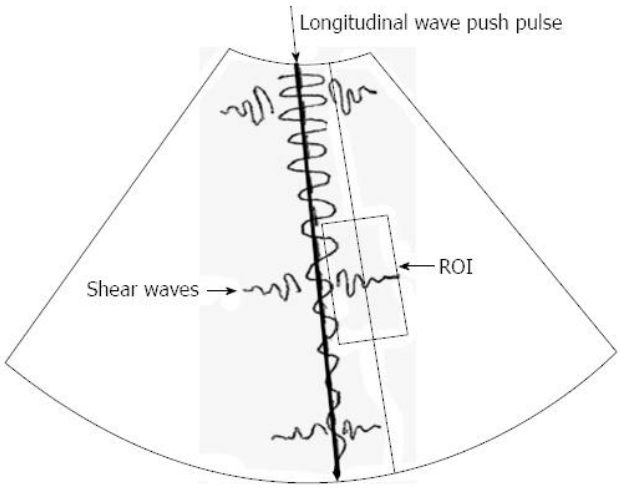
An increasing number of point SWE systems have been developed by different manufacturers, including Virtual Touch Tissue Quantification (VTTQ, Siemens), Elast-PQ

(Philips), as well as newer systems from General Electric, Hitachi and Samsung. VTTQ was the pioneer of the point SWE technique, and remains the best validated and most widely utilised tool in clinical practice. It is widely referred to as 'ARFI'; even though all point SWE and 2D-SWE systems utilise the ARFI push pulse technique. For convention, VTTQ will be referred to as 'ARFI' for the remainder of the thesis.

2.6.1 ARFI - Technical Background

The ARFI technique was originally developed by Kathryn Nightingale's team at Duke University in the USA.[99,141–143] As with all point SWE techniques, the system uses a single high intensity ARFI push pulse to induce micron level displacements in a localized area of tissue. As previously detailed, this induces shear waves to propagate in a perpendicular direction away from the region of excitation. The transducer then switches into imaging mode, using higher frequency ultrasound beams and speckle tracking technique to monitor these displacements at multiple pre-determined locations and time intervals surrounding the impulse location. The time-to-peak displacement and shear wave recovery times across the 10 x 5mm measurement ROI are measured (Figure 2.5), and time-of-flight algorithms are then used to calculate the average regional shear wave velocity.[104]

Figure 2.5: Illustration of the ARFI technique. The longitudinal ARFI push pulse generates shear waves, which propagate away from the region of excitation in a perpendicular direction. The average velocity of their propagation across the region of interest (ROI) is then measured. Figure from D'Onofrio et al.[144]



2.6.2 ARFI - Clinical applications / theoretical advantages

ARFI has been evaluated in a wide range of clinical settings, including the evaluation of thyroid nodules, breast, liver and pancreatic lesions, chronic kidney disease and renal allografts. The most established application however remains the non-invasive assessment of liver fibrosis, as ARFI has a number of theoretical advantages over its predecessors, including TE.

The technique is firstly integrated into a conventional ultrasound machine, and B mode imaging is therefore available to assist with ROI placement. This allows operators to ensure they avoid sampling over large vessels, biliary tracts or masses which have the potential to impact on obtained measurements. The ARFI excitation method also allows the targeted assessment of tissue remote from the ultrasound probe. This enables operators to evaluate deep-seated tissue or to selectively interrogate different parts of the liver; which may be useful in cases of inhomogeneous fibrosis. As previously mentioned, ARFI is hypothesized to be less operator dependent than alternative excitation methods. And on a practical level, ARFI measurements can be acquired by a conventional ultrasound machine rather than requiring a dedicated liver elastography device; which has positive implications for both cost and availability. This allows the integration of elasticity measurements into a conventional ultrasound examination, which may have practical benefits in some patients (e.g. monitoring the progress of cirrhotic patients, whilst they undergoing routine HCC surveillance) and also allows operators to recognize clinical confounders which may affect SWE readings.

2.6.3 ARFI – Liver fibrosis assessment

ARFI has shown high accuracy in assessing liver fibrosis in a number of meta-analyses, which draw from studies performed almost exclusively in Europe and Asia.[15,145] This includes a meta-analysis by Bota *et al.*, which included 13 studies covering 1163 patients with a mixture of CLD aetiologies.[15] They reported an overall sensitivity / specificity of 0.87 / 0.87 for the diagnosis of cirrhosis, and 0.74 / 0.83 for the diagnosis of significant fibrosis. The finding of lower accuracy in diagnosing significant fibrosis (≥F2) is a common finding from most studies, and ARFI's primary utility is therefore in evaluating for F3 or F4 disease; but particularly excluding the presence of cirrhosis.[35] ARFI has naturally been

compared to TE in a number of head-to-head studies, and the accuracy of the two systems has been consistently found to be equivalent.[14,15,22,47,114,146–148]

ARFI has also been specifically validated in context of all major CLD aetiologies. It has had greatest validation in the setting of HCV, with a meta-analysis by Friedrich-Rust *et al.* showing an AUROC of 0.87, 0.91 and 0.93 at detecting ≥F2, ≥F3 and F4 disease respectively.[145] A number of studies and meta-analyses have also demonstrated high accuracy in the assessment of HBV,[145,149]although accuracy levels were less than observed with HCV.[13] The technology has shown slightly lower performance in the setting of NAFLD,[24,148,150,151] with a recent meta-analysis of 723 patients by Liu *et al.* showing an overall sensitivity and specificity of 80.2% and 85.2% in the detection of significant fibrosis.[152] Although more limited, there is also some evidence supporting ARFI's use in alcoholic CLD,[153,154] Primary Biliary Cholangitis,[155] Autoimmune Hepatitis,[156] and biliary atresia.[157]

ARFI has demonstrated high reproducibility in regards to both intra and inter-operator reliability. Intra-operator reliability estimates have ranged between an ICC of 0.84 and 0.96,[34,158,159] whilst inter-operator reliability was found to be nearly as high with an ICC range of 0.81 to 0.93.[158–161]

ARFI also has the benefit of high acquisition success rates. Reliable ARFI measurements can be obtained in over 90 – 95% of patients evaluated,[11,147,162] with overall measurement success rates being relatively higher than TE.[15,23]

2.6.4 ARFI - Cirrhosis severity

ARFI has also been trialed in the assessment of cirrhosis severity in a number of trials, which have shown variable and often conflicting results.

In regards to hepatic reserve, Bota *et al.* found ARFI LSM to have only a weak correlation with Child-Pugh score ($r=0.264$, $p<0.001$) and MELD score ($r=0.194$, $p=0.005$) amongst 211 patients with established cirrhosis.[44] They also found a statistically significant, albeit weak correlation with biochemical parameters including bilirubin ($r=0.271$, $p<0.001$), albumin ($r=-0.270$, $p<0.001$) and prothrombin time ($r=0.196$, $p=0.006$). Vermehren *et al.*

found a similarly weak relationship with cirrhosis severity scores, with ARFI LSM having an AUROC of 0.69 and 0.69 for identifying patients with a MELD score of ≥15 and Child-Pugh B/C cirrhosis respectively.[45]

Studies evaluating ARFI LSM in the setting of portal hypertension, however, have shown more promising results. Salzl *et al*. found ARFI LSM to have a strong correlation with HVPG (r=0.646, p<0.001) amongst 88 patients with established cirrhosis. Whilst predictive utility remained slightly less than TE, ARFI nonetheless showed good accuracy (AUROC 0.855) and good sensitivity and specificity (71.4% and 87.5%, cut-off = 2.58m/s) at diagnosing clinically significant portal hypertension (i.e. HVPG ≥10mmHg).[37] And even higher levels of accuracy have been reported by Attia *et al.*, who found ARFI LSM to have an AUROC of 0.93 and 0.87 for diagnosing HVPG >10mmHg and >12mmHg, respectively.[38]

To further complicate matters, there have been conflicting findings from the numerous studies evaluating ARFI LSM in the prediction of oesophageal varices. Morishita *et al*. found ARFI LSM to have an AUROC of 0.89 (sensitivity = 83%, specificity = 76%, cut-off = 2.05) in identifying patients with any oesophageal varices.[39] And these promising results have been echoed by further groups in Asia.[40,41] Results from two Romania studies, however, lie in direct opposition. Bota *et al*. found ARFI LSM To have poor utility in identifying patients with grade 2 or greater oesophageal varices (AUROC 0.596). And using a cut-off of 2.25m/s they achieved a sensitivity of 93.4% in detecting grade 2 varices, but this came at the expense of a specificity of 28.9%.[163] Similarly poor results were reported by Sirli *et al*. who found no significant difference in the ARFI LSM values between patients with no / grade 1 varices compared to those with grade 2/3 (LSM = 2.73 vs. 2.80m/s, p=0.49).[43] And similarly poor results have been reported by other groups.[42,45] A recent meta-analysis found LSM to have a combined sensitivity of 0.83 (95%CI: 0.78 – 0.87) and a specificity of 0.66 (95%CI: 0.60 – 0.72), however suggested ARFI spleen stiffness measurements showed higher performance in predicting the presence of oesophageal varices.[164]

The role and utility of ARFI in the assessment of cirrhosis therefore remains under debate, and has been identified as an area requiring further research.[165]

2.6.5 Scan factors affecting ARFI performance

There is increasing awareness that a number of parameters can impact on the reliability of obtained ARFI LSMs, which include both patient factors and acquisition technique.

Acquisition parameters

The location of the ARFI measurement within the liver has shown to impact on both LSM velocity but also the measurement reliability. Measurements acquired from the left hepatic lobe have shown to be less reliable and to generate higher SW speeds than the right, which has been attributable to greater cardiac motion effects.[82,83] Segments V, VII or VIII are believed to generate the most accurate results, and are now the recommended target in elastography guidelines.[35]

Measurement depth in the liver has also shown to impact on ARFI performance, with reliability reducing at both extremes. In phantom models, ARFI reliability reduces once measurement depth exceeds 6 – 8cm,[21] with similar findings also being found in humans once the liver capsule to ROI depth exceeded 5.5cm.[166] This is hypothesized to be due to increasing attenuation of the ARFI push pulse and therefore lower acoustic transfer. Conversely, measurements were also found to be associated with lower ARFI accuracy when taken within 1 – 2cm from the liver capsule.[22] This has been attributed to a band of physiologic fibrosis resulting in elevated, non-representative SW velocities and guidelines now recommend acquiring measurements a minimum of 1 – 2 cm and a maximum of 6cm from the liver capsule.[35,93]

Breathing is also thought to impact on ARFI reliability, either as a result of motion artefact or secondary to the valsalva effects of deep inspiration; which may in turn increase hepatic venous pressure and thereby liver stiffness.[94,95,104] Guidelines now recommend ARFI measurements be taken during a light breath hold.[35,93] Finally, ARFI LSM values have been shown to increase for 120 – 180 minutes post-prandially as a result of increased splanchnic blood flow.[19,20] ARFI measurements are therefore recommended to be acquired in a fasting state.

Operator Training

Unlike TE, there is limited information surrounding the impact of operator training and experience on ARFI performance, with only two small studies evaluating the question to our knowledge. The first by Ferraioli *et al.* evaluated 92 healthy volunteers with ARFI by an expert and novice operator over two training periods; assessing intra-observer and inter-observer agreement at these two time points.[33] They found improved measurement reproducibility and inter-operator agreement during the second training period, particularly for the novice operator, and therefore concluded that ARFI was dependent on operator training. The second study by Boursier *et al.* involved 101 patients with CLD, who were also scanned with ARFI by a single operator pair consisting of an expert and novice.[34] They found no difference in accuracy between the two operators and concluded that no training effect existed for ARFI.

The conflicting findings prevent any conclusion being drawn regarding the impact of operator training on ARFI performance. The two small studies also compared performance between a single select pair of expert and novice operators, and it is therefore to extrapolate any differences in operator performance to operators more broadly. Neither study assessed longitudinal ARFI performance over time, with Ferraioli assessing ARFI reproducibility at only two arbitrary time points. Hence if a training effect does exist, its duration cannot be determined.

Reflecting these unknowns, there remains no recommendation regarding a minimum operator training requirement for ARFI in the international elastography guidelines.[35,93] Operators are currently deemed competent to acquire ARFI LSMs from the outset, without any period of probation or supervision being needed. This is in contrast to the well-established guidelines in place for TE, in which operators are required to complete 100 scans for basic competency and 500 scans to be deemed an expert.[3512] Further evaluation regarding the impact of individual operator experience on ARFI performance is therefore seems prudent.

2.6.6 Patient factors affecting ARFI performance

Necroinflammatory change

Necroinflammatory change, as indicated by histopathology or an elevated ALT level, has been consistently associated with higher ARFI velocities in a number of studies.[16–18] This increase in ARFI velocity has been independent of fibrosis levels, and has been attributable to the inflammatory infiltrate increasing liver stiffness. Inflammatory activity has also shown to reduce ARFI accuracy,[17,167] particularly once the ALT level was over 5 times the upper limit of normal.[17]

Ascites is felt not to significantly affect ARFI reliability, which is one of the purported advantages of ARFI over TE. The additional fluid interface is theorized not to cause significant attenuation of the ARFI push pulse, allowing the underlying liver to be interrogated without impedance.[27] Whilst confirmation in clinical trials has been limited, Bota *et al*. found ARFI to have good accuracy in discriminating between cirrhotic and non-cirrhotic ascites – lending support to the above claims.[168] As with TE, hepatic congestion from right heart failure has also been linked with increased liver stiffness values and therefore lower ARFI accuracy.[19]

Obesity

Body habitus is also thought likely to impact on ARFI reliability, with a number of studies linking obesity with a reduction in a wide range of performance parameters. Increased BMI has been linked with a higher failure rates of IQR/Median reliability criteria,[11,23] with unreliable measurements reported in 48.8% of obese patients compared to 14.5% of non-obese patients by Cassinotto *et al*. ($p<0.0001$). Intra & inter-observer reliability also appears to be slightly lower amongst patients with a normal BMI (ICC = 0.91 and 0.82 respectively) compared to those who are overweight or obese (i.e. BMI ≥25kg/m^2, ICC = 0.88 and 0.79),[158] And a number of studies have reported ARFI accuracy to also be reduced in the setting of obesity.[23,24,169] The most cautionary findings come from the Bordeaux group who analysed ARFI performance amongst 321 patients with mixed aetiology CLD. They found accuracy to be markedly reduced amongst obese patients (i.e. BMI ≥30kg/m^2 vs. <30kg/m^2) both in diagnosing cirrhosis (AUROC = 0.63 vs. 0.92, $p=0.0002$) and severe fibrosis (AUROC = 0.63 vs. 0.91, $p<0.0001$). The mechanism

underlying obesity's likely negative association with ARFI reliability has been theorised to involve attenuation of either the ultrasound push pulse or tracking beams from increased depths of subcutaneous adipose tissue,[148,170] however this hypothesis has not been further evaluated to our knowledge.

The relationship between BMI and ARFI reliability has not been universally reported by all studies however.[28,47] Attia *et al.* analysed ARFI performance amongst 97 overweight or obese patients and found very high accuracy in both the overweight (BMI = 25 – 30kg/m^2, AUROC = 0.97) and obese subcohorts (BMI>30kg/m^2 AUROC = 0.94).[28] BMI was also found not to be associated with increased rates or discordance between ARFI and liver biopsy in multivariate analyses (p=0.245) and the study therefore concluded that "BMI does not influence the staging of liver fibrosis using acoustic radiation force impulse imaging elastography in obese patients". High accuracy has also been reported for a number of studies assessing the performance of ARFI in the setting of NAFLD.[148,150,151] These cohorts include high rates of overweight and obese patients, with high accuracy (Az = 0.899) even being observed by Guzman-Aroca *et al.* amongst 32 morbidly obese patients assessed for NAFLD vs. NASH prior to bariatric surgery.[29] These encouraging results are in stark contrast to the cautionary findings of some studies mentioned above. Therefore whilst body habitus appears to negatively impact ARFI performance, the magnitude of its impact, and the importance of obesity as a confounding factor remains uncertain.

Hepatosteatosis

The impact of hepatosteatosis on ARFI liver fibrosis assessment remains less well defined. There is some evidence that hepatosteatosis causes a reduction in LSM values independent of fibrosis levels, which is thought to reflect softening of the liver from increasing deposition of fat.[32,148] For example, Yoneda found steatosis to be associated with reducing LSM values (p=0.03) amongst a group of 44 patients with NAFLD.[148] This relationship has not been consistently observed by all studies however,[24,171,172] with Lupsor *et al.* finding ARFI LSM to correlate with fibrosis (r=0.717, p<0.0001), necroinflammatory activity (r=0.328, p=0.014) but not with steatosis (r=0.122, p=0.321) amongst a group of 112 patients with HCV [172]. Similarly Bota *et al.* found steatosis to have no correlation (r=0.03, p=0.72) amongst 82 patients with HBV and HCV.[171] To further cloud the picture, animal

experiments have paradoxically linked hepatosteatosis with higher SW velocities.[173,174] Guzman Aroca *et al*. assessed ARFI LSMs amongst chickens fed a standard vs. hyperlipidemic diet and found ARFI velocities to be significantly higher in the hyperlipidemic group (SWV = 1.91 vs. 0.94m/s, p<0.001) and ARFI LSMs to have a strong correlation with steatosis severity (r=0.85, p<0.001).[161]

Whilst steatosis is widely recorded as a patient characteristic, analyses regarding its impact on ARFI performance measures is surprisingly limited. Bota *et al*. assessed the impact of hepatosteatosis amongst a group of 82 patients with HBV or HCV. [31] They found patients with moderate to severe steatosis (Hepburn IV and V) to have a poorer correlation between ARFI LSM and histopathology than patients with no or mild steatosis (r = 0.223 vs. r=0.535) and also to have higher failure of the IQR and SR criteria (33.3% vs. 10% failure rates respectively). They therefore concluded that steatosis was an important factor impacting on ARFI accuracy. The study did not control for BMI, however, which is the major limitation of the analyses. Another study by Attia *et al*. looked at whether steatosis was independently associated with increased rates of discordance between ARFI LSM and biopsy in multi-regression analyses. They didn't observe a relationship between steatosis and accuracy amongst overweight (p=0.124) or obese patients (p=0.100), however this may have been attributable to power; with only 6 of the 87 patients having ARFI values which were discordant with biopsy. There is also some evidence from NAFLD rat models that severe steatosis is associated with reduced ARFI accuracy amongst rodents with lower fibrosis levels (i.e. F0/F1); however whether this is finding is transferrable to humans is unclear.[173].

2.6.7 ARFI - Reliability assessment

Adopting lessons learnt with TE,[126,127] ARFI now routinely uses the IQR/Median as a surrogate indicator of measurement reliability. The IQR/Median ratio reflects the internal consistency of measurements obtained within a patient, with high values indicating greater spread between the ten LSM readings. The approach relies on the premise that obtaining inconsistent LSM values within a patient (i.e. high IQR/Median) increases the likelihood of the overall LSM result being unreliable. The approach has been validated in the setting of ARFI by a number of groups, with Bota *et al*. finding ARFI accuracy to be reduced amongst

scans with an IQR >0.30 or SR <60% (AUROC = 0.268 vs. 0.722), whilst Goertz *et al.* found similar findings but using a SD/Mean cut-off of >0.30.[36,175] The approach is however imperfect, and patients who fulfilled the criteria still had issues with reduced sensitivity at identifying lower levels of fibrosis (i.e. sensitivity of 72.5% and NPV of 58.8% at the F1/F2 cut-off).[36]

Borrowing from findings with TE,[128] there is also the suggestion that the IQR/Median criteria may only be applicable at a minimum ARFI LSM threshold. The elastography guidelines from Barr *et al.* suggest the IQR/Median ratio may only be relevant amongst ARFI readings with a LSM >1.50m/s, however this approach is yet to be validated.[140]

No further methods have been validated or are in broad clinical use for the assessment of ARFI reliability. This is particularly unfortunate for point SWE, as the absence of an elastogram prevents operators from qualitatively assessing the quality of shear wave generation and propagation. New reliability assessment strategies which may allow clinicians to better gauge the validity of obtained ARFI measurements are therefore needed.

2.7 Two Dimensional Shear Wave Elastography (2D-SWE)

2.7.1 2D-SWE – technological principles

2D-SWE is the most technologically advanced ultrasound elastography technique in current clinical use, and represents an extension of the point SWE technique. Like point SWE, 2D-SWE uses ARFI push pulses to excite a region of liver parenchyma and then tracks generated shear waves using ultrasound beams and the speckle tracking technique. 2D-SWE however differs from point SWE by the use of sequential ultrasound push pulses to stimulate multiple different locations in the liver. Shear wave arrival times are then monitored at numerous points surrounding these multiple areas of excitation, allowing the elasticity profile through a larger section of liver to be re-constructed.

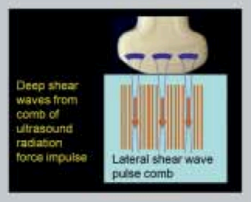
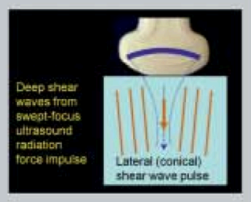
2D-SWE can allow either qualitative or quantitative assessment of tissue elasticity. Qualitative analyses can be achieved through the reconstruction of a two-dimensional shear wave velocity map, which can then be superimposed onto a B-mode image for

review. This can be evaluated as a static image (i.e. by summating a small number of push pulse sequences), however some systems have the capacity to generate a continuously refreshing image by repeatedly performing the push pulse sequence. This 'continuous mode' however requires a mandatory minimum 'cooling time' between each ultrasound push pulse sequence to allow tissue recovery and prevent overheating. This in turn limits imaging frame-rate, and therefore the resolution of the continuous mode imaging. 2D-SWE also allows quantitative assessment by allowing an operator to place a measurement ROI in an area of interest on the elasticity map. The system then calculates the shear wave velocity parameters for this area; which usually includes the median, mean, maximum and standard deviation.

2D-SWE has a number of theoretical advantages over other ultrasound SWE systems, including ARFI. It firstly allows both qualitative and quantitative assessment of liver elasticity, which is not provided by other ultrasound based systems to date.[176,177] The ability to visualize shear waves firstly allows operators to assess spatial variations in liver stiffness, and thereby ensure that the samples are representative of the overall elasticity profile.[178] The shear wave maps allow operators to recognize areas of artefact, which should be avoided when positioning the ROI.[176] And the ability to visualize shear wave propagation may provide operators with an indication of the quality and therefore reliability of obtained measurements.

A number of different 2D-SWE systems have been developed, which includes Supersonic Shear Imaging (Aixplorer, Supersonic Imagine), Virtual Touch Imaging Quantification (VTIQ, Siemens, Germany), systems by both Phillips and General Electric as well as a new 2D-SWE system by Toshiba Medical Corporation (Tochigi, Japan). Whilst these all share technological principles, there are a number of important technical differences between the systems. The pre-eminent difference is the method used to sequentially excite the liver, which is summarized in the Figure 2.6.[35] The most common method, used by both the Toshiba and Phillips systems, generates ARFI push pulses at multiple different depths within a single axial plane. The General Electric system also excites tissue at different depths in an axial plane, however does this along multiple parallel lines creating a 'comb' of excitation. And finally, Supersonic Shear Imaging sweeps sequential ARFI push pulses across a section of tissue to create a cylinder of shear waves known as a 'mach cone'.

Figure 2.6: Summary of the 2D-SWE systems in clinical development, including the different methods of excitation employed by the different devices. Image from the EFSUMB elastography guideline.[35]

ultrasound induced – radiation force impulses focused at various depths	shear wave speed[5]	quantitative	single image within a colour box	Siemens	
			image within a colour box, running refresh	Toshiba Philips Mindray Zonare	
ultrasound induced – radiation force down multiple simultaneous lines in a "comb push" combined with directional filtering	shear wave speed[5]	quantitative	single image within a colour box	GE	
ultrasound induced – radiation force focus swept over depth faster than shear wave speed to create a Mach cone	shear wave speed[5]	quantitative	image within a colour box, refreshed at up to several per second[3]	SuperSonic Imagine	

There is also some variation in the image processing and display modes employed by the different systems. Whilst all of the devices produce a static elasticity map from which quantitative measurements can be taken, the continuous mode is only provided by some vendors (including both SSI and Toshiba 2D-SWE). There are also differences in the way regional shear wave characteristics are visually presented. Whilst most 2D-SWE devices provide a colour map of shear wave velocities superimposed on a grayscale B-mode image for anatomic reference, some vendors provide additional imaging interfaces (discussed in more detail below).

Finally, whilst all 2D-SWE systems assess the same underlying physiologic property (i.e. tissue elasticity), the shear wave velocity obtained by the respective systems has shown to be non-equivalent. Hall *et al.* reported an approximately 12% variance in the average shear wave values obtained by different vendors in the same patient.[179] The differences are attributable to a number of system variables. In addition to software / processing differences, the 2D-SWE systems use variable ARFI push pulse frequencies which impact on the velocity of the created shear waves. Efforts are currently underway to standardize

the shear wave velocities between systems. Nonetheless the current differences prevent the direct application of shear wave cut-offs and technical guidelines between the different 2D-SWE systems.

2.7.2 Supersonic Shear Imaging (SSI)

Supersonic Shear Imaging (SSI, Aix-en-Provence, France) is the primary 2D-SWE technique in current clinical use, and is often referred to by its trade name 'Aixplorer'.[176,177,180,181] Whilst SSI remains less evaluated than either transient elastography or ARFI, it represents the most validated 2D-SWE technique and is therefore the reference for similar technologies in clinical development.

As previously mentioned, SSI utilises sequential ultrasound push pulses to stimulate different depths in the liver. The excitations are however performed at a speed greater than subsequent shear wave propagation, and thereby produce a three-dimensional cylindrical wave front known as a 'mach cone' (Figure 2.7).[180] The other main feature of SSI is its use of ultrafast imaging to track subsequent shear wave propagation. This uses frame rates of up to 5,000 images per second, and allows SSI to assess the shear wave propagation of the entire mach cone during a single push pulse sequence. This not only facilitates continuous, real-time image but also improves image quality (Figure 2.8).[100]

Figure 2.7: Supersonic Shear Imaging and the formation of a 'mach cone' through the use of multiple sequential excitations. Image from Gennisson et al.[100]

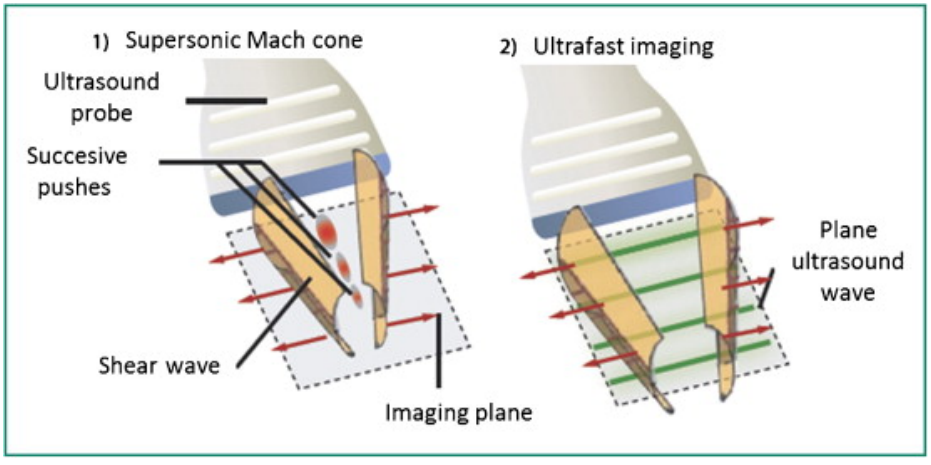
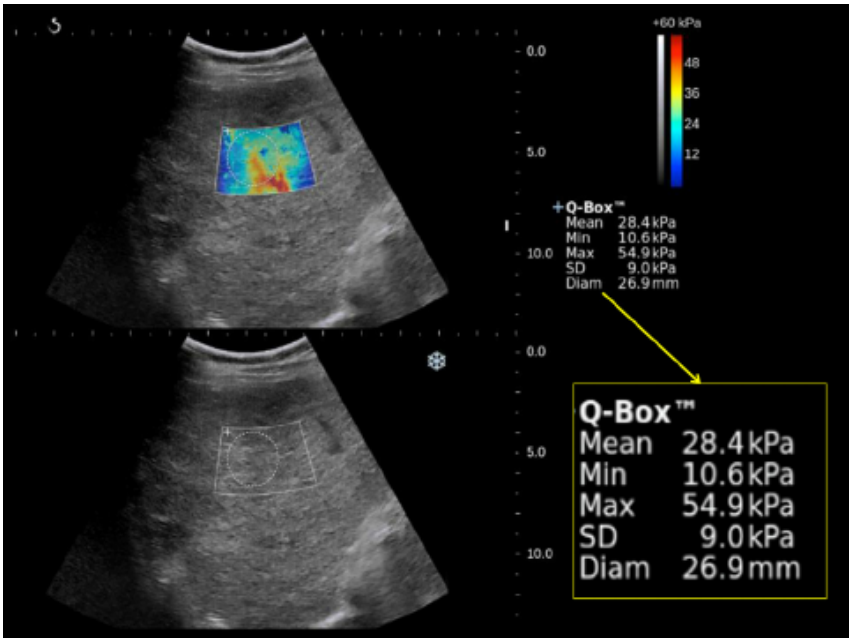


Figure 2.8: Screenshot of a SSI acquisition. The top picture represents the shear wave velocity map superimposed on a B-mode image; higher shear velocities being represented in orange / red and lower velocities shown in blue / green. A description of the shear wave velocity properties within the adjustable region of interest (i.e. mean, minimum, maximum and standard deviation) are then displayed. Figure from Piscaglia et al[182].



SSI has been shown to have high accuracy in the assessment of liver fibrosis; multiple studies finding it to be equivalent if not superior to TE and ARFI in the assessment of liver fibrosis. [15,46,47,178,24,183,184] In a head-to-head study involving 349 patients, Cassinotto *et al.* found SSI to have an AUROC of 0.89, 0.88 and 0.93 at discriminating the F1/F2, F2/F3 and F3/F4 cut-offs, compared to 0.86, 0.84 and 0.90 for TE and 0.84, 0.81 and 0.90 for ARFI.[46] SSI has also shown high intra and inter-observer reliability[185] as well as high internal measurement reproducibility; with one study suggesting as few as three measurements may be required in the assessment of liver fibrosis.[186]

Similar reliability issues have however been observed in SSI as in ARFI,[35] with SSI performance shown to be affected by measurement depth[187] as well as obesity.[24,47,183,188]

2.7.3 Toshiba 2D-SWE

A new 2D-SWE technique has been developed by Toshiba Medical Systems Corporation (Tochigi, Japan), which is currently in the early phases of clinical development.[189] As previously mentioned, Toshiba 2D-SWE sequentially pulses tissue in a single axial plane, and then assesses regional shear wave propagation using tracking ultrasound beams and the speckle tracking technique. These shear wave display maps can be viewed in 'continuous' mode, which involves continuously refreshing images following each push pulse sequence. A summation of multiple ultrasound push pulse sequences can however be used to produce a higher quality 'single shot' image, which can then be used for quantitative analysis. The procedure underlying the acquisition of a quantitative measurement is further outlined in Chapter 3.2.

The technique is novel in the provision of two display maps for the 'single shot' scan, which provide different visual representations of the liver's shear wave profile. The first is the 'Speed Smart Map', which provides a color representation of regional shear wave velocities within a section of liver. This is similar to other 2D-SWE vendors including SSI, with areas of liver with high shear wave properties displayed as orange / red and areas of slower shear wave velocity in blue / green. Areas of suboptimal shear wave propagation characteristics are displayed in black to forewarn operators of potential areas of artefact. The second 'Propagation Map' is however unique to the Toshiba system, and uses contour lines to depict the arrival time of shear waves at different points in the tissue.[189] Contour lines which lie close together therefore represent slow shear wave propagation, whilst those lying far apart reflect higher velocity waves. The shape of the wave also help visualize the uniformity of elasticity properties within a tissue. Whilst contour lines are parallel in completely uniform / isotropic tissue, the presence of irregular wavefronts indicates the heterogeneity of the tissue's underlying elasticity characteristics. These two display maps in theory provide complementary information regarding shear wave propagation. The additional information provided by the Propagation Map is designed to allow operators to better assess the suitability of a 'single shot' acquisition for quantitative analysis and to optimize ROI positioning.[189]

Toshiba 2D-SWE is currently in the early clinical stages of development. As a result, there is minimal published data on Toshiba 2D-SWE and detailed acquisition guidelines for the technique are yet to be established. Whilst literature for SSI and ARFI provide a starting

point, the multiple aforementioned technical differences prevent their direct application for the new system.

Chapter 3. Methods

The two shear wave technologies; ARFI (Siemens) and 2D-SWE (Toshiba) were assessed amongst two separate patient cohorts with differing methodologies. These are described separately in Chapters 3.1 and 3.2 for ARFI and 2D-SWE systems, respectively.

3.1 Acoustic Radiation Force Impulse elastography (ARFI)

3.1.1 Ethics

The research project was approved by the Melbourne Health Office of the Human Research Ethics Committee (HREC Reference Number = 2012.259). The project was approved as a 'low risk' quality assurance exercise and the need for patient consent was therefore waived.

3.1.2 Patient cohort

The greater ARFI cohort incorporated all patients with CLD who were scanned with ARFI at the Royal Melbourne Hospital as part of clinical liver fibrosis assessment. The cohort included patients with established diffuse chronic liver disease as well as those being investigated for suspected CLD pathology. Patients were required to be over 18 years of age for review. There were no additional exclusion criteria relating to patient demographics or CLD details, with the cohort consisting of wide ranging CLD aetiologies and severities, including those with very mild disease through to cirrhosis. ARFI was performed as part of routine patient care, with ARFI representing the institution's first line elastography tool used for liver fibrosis evaluation. The majority of patients were internally referred from doctors working within Melbourne Health, predominantly from the gastroenterology and infectious diseases units. All attempted ARFI studies were reviewed, including those in which measurements were unable to be successfully obtained. All ARFI scans performed from ARFI's clinical introduction in August 2012 until November 2014 were reviewed, which encompassed 934 patients in total.

The rationale and implications of the deliberately broad inclusion criteria is further discussed in Chapter 5.5.

3.1.3 Patient clinical information

Clinical information was obtained retrospectively from hospital information services. Sources of medical information accessed included the hospital's paper medical records, radiology information system (Synapse), pathology results service, online medical records and endoscopy results service.

Information collected included the patient's age at the time of ARFI examination, gender and relevant past medical history which may have impacted on cirrhosis scores; including Gilbert Syndrome and non-cirrhotic causes of coagulopathy, thrombocytopaenia or ascites. The aetiology of chronic liver disease was based on physician assessment as documented in medical records. In patients with multifactorial liver disease, all contributory pathologies were recorded and counted individually. The height and weight was routinely measured at the time of ARFI for all patients scanned consecutively from April 2014 onwards. This was used for the calculation of Body Mass Index using standard formula: BMI = weight / height^2 (kg / m^2). As per world health organization (WHO) recommendations, overweight was defined as a BMI $\geq$25kg/m^2 and obese as BMI $\geq$30kg/m^2.[190]

Results of blood tests performed closest to the date of ARFI was recorded for all patients. This included liver function tests (LFTs), urea electrolytes and creatinine (UEC), full blood examination (FBE) and pro-thrombin time international normalized ratio (INR). The elapsed period in days between each blood test and the ARFI examination was calculated. A maximum three months cut-off between blood tests and ARFI was applied, to help ensure pathology results provided a true representation of liver disease status at the time of ARFI assessment. The pathology results were used for the calculation of cirrhosis prognostic scoring systems (described below)

Disease activity factors were recorded for patients, including the presence of ongoing alcohol intake in alcoholic hepatitis and viral load and antibody status in patients with viral hepatitis. This information was however incomplete and deemed of insufficient reliability to warrant analysis. Only a small proportion of patients (14%) in the ARFI cohort had undertaken a Fibroscan™ assessment, many of which were performed temporally remote from ARFI. Due to the limited power and high risk of selection bias, comparative analyses between ARFI and transient elastography have not been performed.

3.1.4 Cirrhosis information

To be included in the cirrhotic subcohort, patients required the clinical diagnosis of cirrhosis which was not based on elastography parameters. This included the presence of F4 disease on histopathology (in those patients with prior liver biopsy results available), morphological changes of cirrhosis on anatomic imaging, characteristic biochemistry and pathology abnormalities or the clinical development of cirrhotic or portal hypertensive complications.[191]

The Model for End-stage Liver Disease (MELD) Score was calculated for cirrhotic patients using the formula listed below (Figure 3.1). MELD is a prospectively validated scoring system developed for the prediction of three months mortality in patients with CLD, which is now widely employed in the allocation of liver transplants.[192,193] The modified Child-Pugh Score was also calculated for cirrhotic patients (formulae outlined in Figure 3.1). This is also a well-validated tool which was originally developed for the assessment of peri-operative mortality in patients with CLD, but is now used more broadly to assess the general prognosis of cirrhotic patients.[194] Patients in whom blood tests results were performed over three months distant to ARFI or who had inadequate information to generate a valid MELD or Child-Pugh Score were deemed to have missing data. Patients on anticoagulants were also excluded from analyses.

Figure 3.1 : The MELD score and Child-Pugh score were calculated using standard formulae as outlined below.[192,194]

MELD Score: $(0.957 \times \ln(\text{Creatinine}) + 0.378 \times \ln(\text{Bilirubin}) + 1.12 \times \ln(\text{INR}) + 0.643) \times 10$

Child-Pugh Score:

- Bilirubin:
  - 1: <34.2 µmol/L
  - 2: 34.2-51.3 µmol/L
  - 3: >51.3 µmol/L
- Albumin:
  - 1: >35 g/L
  - 2: 28 - 35 g/L
  - 3: <28 g/L
- INR:

	1:	<1.7
	2:	1.7-2.2
	3:	>2.2
- Ascites:		
	1:	Absent
	2:	Mild
	3:	Moderate – Severe
- Encephalopathy:		
	1:	No encephalopathy
	2:	Grade I – II
	3:	Grade III - IV
Child-Pugh Grade:		
	A	Score: 5 – 6
	B	Score: 7 – 9
	C	Score: 10 – 15

The prediction of varices was assessed in a subgroup of patients with cirrhosis who completed upper gastrointenstinal endoscopy, primarily for the surveillance or treatment of oesophageal varices. Of these 122 patients, 59 patients completed gastroscopy within six months of ARFI and were included analyses. Six months was felt to be an appropriate cut-off to ensure the gastroscopy findings were temporally reflective of the cirrhosis severity at the time of ARFI examination. The presence/absence of oesophageal and gastric varices, as well as their grade, was recorded for all patients. Those patients with a history of oesophageal varices which had been successfully ablated at the time of follow-up gastroscopy were considered to have varices; the recorded size being that prior to banding. A preceding history of variceal bleeding and need for blood transfusion was also recorded for assessment of cirrhotic decompensation.

The presence of ascites was diagnosed on either physician assessment (i.e. physical examination as documented in medical records) or on radiological studies (primarily ultrasound). Patients with a history of ascites which had been successfully treated at the time of ARFI examination were deemed to have ascites. The grade of ascites (mild, moderate or severe) as well as the need for prior large volume paracentesis was also recorded.

The presence and severity of encephalopathy was also based on physician assessment, as documented in medical records. The severity of encephalopathy was clinically graded from 1 to 4, as per the widely accepted West Haven Criteria (Table 3.1).[195] The need for treatment with either lactulose or Rifaximin was also recorded. For those patients on lactulose or Rifaximin, encephalopathy severity was as documented prior to the institution of therapy.

Table 3.1: West-Haven criteria for scoring hepatic encephalopathy severity.[195]

Stage	Clinical Manifestations
Minimal	Abnormal results on psychometric or neuropsychological testing without clinical manifestations
1	Changes in behavior, mild confusion, slurred speech, disordered sleep
2	Lethargy, moderate confusion
3	Marked confusion (stupor), incoherent speech, sleeping but arousable
4	Coma, unresponsive to pain

A current or prior history of hepatocellular carcinoma (HCC) was also recorded. Diagnoses were based on radiological features or histopathology (were available), as per standard consensus guidelines.[196,197] Additional cirrhotic complications recorded included the hepatorenal syndrome (as made on clinician assessment) and hepatopulmonary syndrome.

3.1.5 ARFI acquisition

ARFI was performed with the Acuson S2000 ultrasound system (Siemens, Mountain View, CA, USA); a general purpose machine used for wide-ranging clinical ultrasound applications (Figure 3.2). Liver stiffness measurements were acquired using the Virtual Touch Tissue Quantification (VTTQ) imaging application. The lower-frequency 4C1 convex transducer probe (1 – 4.5 MHz) was used to acquire all measurements.

Figure 3.2. The Acuson S2000 ultrasound machine from Siemens, used for the acquisition of ARFI readings. (Image reproduced from Siemens Website)

ARFI measurements were acquired following traditional B-mode ultrasound examination of the liver. Patients were assessed in the supine position with right arm abducted, to help maximize the intercostal acoustic window. Measurements were acquired following a minimum six hours fast to minimise post-prandial hepatic congestion. Measurements were acquired with breathing held in light suspension, making sure to avoid deep breath-holding which can increase hepatic congestion from valsalva effect. Operators were instructed to maintain firm contact between the probe and skin, but to avoid heavy pressure which may compress the liver and thereby elevate liver stiffness measurement (LSM). The exact pressure applied to the skin (i.e. in Newtons) was not formally assessed/regulated in this study. Measurements were acquired via an intercostal approach, with subcostal readings obtained in limited cases where intercostal measurements were unable to be reliably acquired (i.e. due to narrow intercostal spaces). Elastography measurements were acquired from the right hepatic lobe, in a region of liver with a good acoustic window. The ultrasound beam was angled perpendicular to the liver capsule to limit refraction of the 'push pulse' (Figure 3.3). Measurements were obtained from non-identical positions in liver parenchyma away from portal tracts or vascular structures. The

mean shear wave velocity was calculated within a rectangular measurement ROI of fixed 5 x 10mm dimensions. ARFI acquisition continued until 10 valid measurements were obtained. Invalid ARFI readings (indicated by an 'X.XX' on acquisition screenshots) were also recorded. These were used for calculation of ARFI success rate (SR = successful ARFI readings / total readings); which was previously utilised as a marker of ARFI reliability.[31] The median velocity of the ten valid measurements was used as the overall ARFI velocity for each measurement set. This was automatically calculated and displayed on the ARFI summary screenshot generated at the conclusion of each ARFI scanning session (Figure 3.4). The IQR/Median ratio was calculated for all obtained ARFI examinations. As per current consensus guidelines, a scan 'failing' reliability criteria was defined as those with an IQR/Median >0.30.[35,93,140] ARFI velocity cut-offs drawn from the meta-analysis of Friedrich Rust *et al.* were used to assign an F score from the median ARFI velocity.[145] These cut-off values were based from eight studies (518 patients), with the cut-offs for significant fibrosis (≥F2), severe fibrosis (≥F3) and cirrhosis (F4) being 1.34, 1.55 and 1.80m/s respectively.

Figure 3.3: Screenshot of an ARFI acquisition, which also demonstrates the measurement of skin-to-liver capsule distance (SLD).

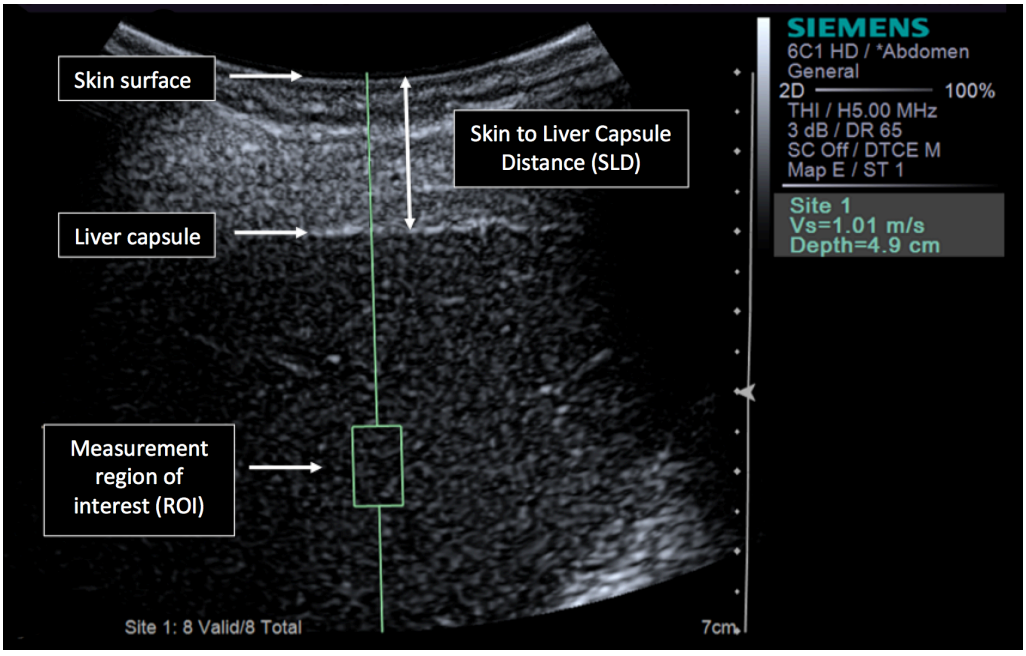
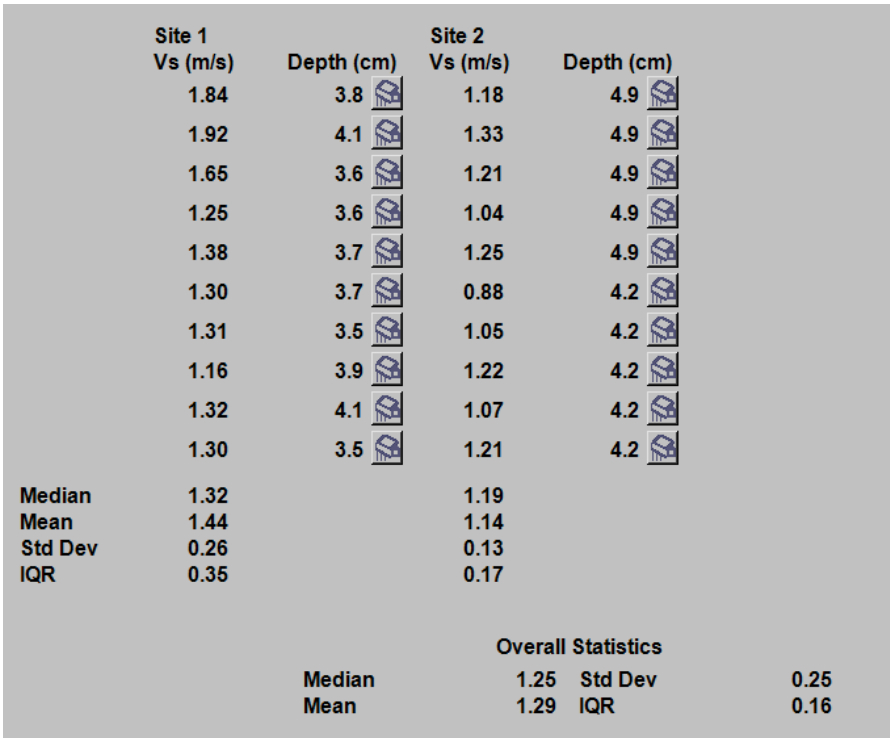


Figure 3.4: Summary screenshot generated at the conclusion of an ARFI session; listing all individual ARFI measurements as well as the overall median velocity and IQR. 'Site 1' and 'Site 2' refer to two measurement sets obtained by different operators.



	Site 1 Vs (m/s)	Depth (cm)	Site 2 Vs (m/s)	Depth (cm)
	1.84	3.8	1.18	4.9
	1.92	4.1	1.33	4.9
	1.65	3.6	1.21	4.9
	1.25	3.6	1.04	4.9
	1.38	3.7	1.25	4.9
	1.30	3.7	0.88	4.2
	1.31	3.5	1.05	4.2
	1.16	3.9	1.22	4.2
	1.32	4.1	1.07	4.2
	1.30	3.5	1.21	4.2
Median	1.32		1.19	
Mean	1.44		1.14	
Std Dev	0.26		0.13	
IQR	0.35		0.17	

Overall Statistics			
Median	1.25	Std Dev	0.25
Mean	1.29	IQR	0.16

The distance from the skin surface to the liver capsule or 'skin-to-liver capsule distance' (SLD) was manually measured on all individual measurements from the 934 patients in the study cohort (Figure 3.3). Measurements were performed by a single operator using the picture archiving and communication service (PACS) ruler function, with measurements recorded to three decimal places. The distance from the liver capsule to the edge of the measurement ROI was also manually measured to three decimals for all acquired measurements. It should be noted that SLD was measured retrospectively from each individual ARFI acquisition screenshot. This is distinct from TE in which SLD is measured separately and prospective to LSMs readings to help assist appropriate probe selection (i.e. M vs. XL probe).

Any technical issues encountered by operators during the ARFI examination was noted. Issues documented included difficulty breath-holding, high lying liver, bowel gas, severe hepatosteatosis, body habitus and generally poor patient compliance.

The presence and severity of hepatosteatosis was graded according to the level of echogenicity and ultrasound beam attenuation seen on B-mode imaging.[30] Steatosis was graded into absent or mild steatosis, or moderate to severe categories using the sonographic features listed below (Table 3.2).

Table 3.2. Grading of hepatosteatosis based on ultrasound features.[30]

Grade of Hepatosteatosis	Ultrasound features
No	Normal echogenicity
Mild	Slight but definite increase in liver echogenicity, with no sonographic features of moderate to severe steatosis (listed below)
Moderate	Moderate or marked diffuse increase in liver echogenicity, with impaired or no visualisation of the portal tracts +/- focal fatty sparing
Severe	Marked diffuse increase in liver echnogenicity, with impaired or no visualisation of the portal tracts and beam attenuation.

3.1.6 Operators

ARFI measurements were acquired by a total of 29 different operators throughout the course of the study. Operators consisted predominantly of sonographers from the hospital's radiology department (n=24), as well as a small number of doctors (n=5). Operators had a broad range of experience in ultrasound, ranging from trainees to a senior radiology consultant with >30 years sonographic experience. Operators completed basic training in ARFI elastography prior to the technology's implementation, but had no other prior point SWE experience. Operator identity was recorded for all scans, with the operator's cumulative ARFI experience calculated for each study performed. The institution's overall scan experience (i.e. cumulative number of patients tested) was also calculated for each ARFI examination.

Patients were routinely assessed by multiple different operators, each of whom acquired 10 valid ARFI LSMs. Operators assessed patients independently and were blinded to each other's results. The majority of patients (n = 640) were routinely tested with two independent operators, however a sub-cohort of 291 patients scanned consecutively from January 2013 until March 2014 were all routinely tested by three operators. To improve standardisation, inter-operator agreement/concordance was only assessed between the first two operators unless stated otherwise (i.e. the third operator values were not included in analyses).

3.1.7 Liver Biopsy

212 patients in the cohort had a liver biopsy performed as part of the clinical management of their CLD. The majority of these biopsies were temporally distant from the ARFI assessment, with the mean elapsed period between ARFI and biopsy being 1,131 days (q1-q3: 131 - 1,990 days). A maximum cut-off of six months between liver biopsy and ARFI was applied to improve the applicability of the biopsy results. Liver biopsies of ≤15mm in length or containing ≤5 portal tracts were also excluded from analyses, as these have been associated with lower accuracy from sampling error.[71,72] Fifty-five patients were ultimately included in the liver biopsy sub-cohort for use in accuracy analyses. All specimens included in the ultimate accuracy sub-cohort were core biopsies performed via a percutaneous technique. The breakdown of patient selection is demonstrated in Figure 3.5.

Recent liver biopsy specimens (performed within six months) were re-reviewed by a single pathologist with a special interest in liver histopathology (MC). The pathologist was blinded to all clinical data and ARFI scores at the time of review. The severity of fibrosis was re-graded using the Metavir scoring system (Table 3.3).[198] This is a five-point scale which grades fibrosis severity from F0 (no fibrosis) through to F4 (cirrhosis). This scoring system was traditionally designed for the assessment of Hepatitis C, however has now been adopted as the most widely utilised classification system in studies assessing patient cohorts encompassing a range of chronic liver disease aetiologies. The Metavir scoring system was chosen in preference to either Scheuer[199] or Ishak[200] due to its wider application in the ARFI literature, with a view to increasing the generalizability of our results.

Figure 3.5: Inclusion criteria of the liver biopsy subcohort.

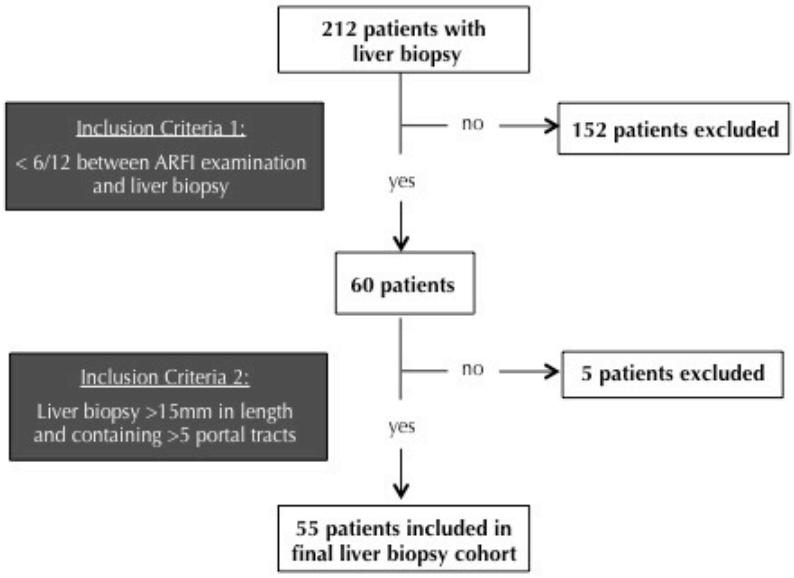


Table 3.3 : Metavir fibrosis score, which grades the severity of liver fibrosis on a five point scale (F0 to F4).[198]

Fibrosis stage	Histopathologic description
F0	No fibrosis
F1	Mild fibrosis – portal fibrosis without septa
F2	Moderate fibrosis – portal fibrosis and few septa
F3	Severe fibrosis – numerous septa without cirrhosis
F4	Cirrhosis

The degree of necroinflammatory change was graded using the Metavir activity score, also known as the 'A score' (Table 3.4).[198] The A score encompasses both piecemeal and focal lobular necrosis, with necroinflammatory activity graded from A0 (no inflammation) to A3 (severe inflammation).

Table 3.4 : Metavir activity score, which grades liver inflammatory change on a four point scale (A0 to A3).[198]

Score	Activity severity
A0	None
A1	Mild
A2	Moderate
A3	Severe

Hepatosteatosis was graded using the Brunt scoring system (Table 3.5); the most widely utilised scale for the grading hepatosteatosis.[201] A visual estimation of the percentage of hepatocytes affected with macrosteatosis was also recorded.

Table 3.5: Brunt grading system for hepatosteatosis.[201]

Brunt Grade	Percent of hepatocytes in the biopsy involved
0	None
1	<33%
2	33-66%
3	>66%

3.1.8 Statistical analyses

Analyses were completed using the Graphpad Prism 8 and Stata 12 statistical programs.

IQR/Median ratio was calculated for all measurement sets, as this has been shown to be an established marker of measurement reliability in ARFI.[31,36] The proportion of ARFI studies failing IQR/Median criteria (i.e. IQR//Median >0.30) was recorded. The impact of factors affecting the consistency of individual measurements (e.g. measurement depth in the liver) was assessed by calculating the deviation of individual measurements in m/s from the set's median.

Inter-operator agreement was assessed using three different approaches. Amongst selected operator pairs who had performed >20 scans together (n=4), an intra-class correlation coefficient (ICC) was calculated. Percentage deviation between the first two operators was calculated as the velocity difference between operators (m/s) divided by their mean ARFI LSM. Inter-operator concordance was defined as operators obtaining ARFI velocities which were in the same or adjacent F score ranges; the rationale for this definition is discussed in Chapter 5.5. Percentage deviation between operators was considered the primary measure of inter-operator agreement, as the lack of an agreed definition for inter-operator concordance in the literature may lessen the robustness of inter-operator concordance as a measure. Inter-operator agreement was assessed between the first two operators only, unless stated otherwise.

Accuracy was assessed using the receiver operator curve analysis (ROC). Sensitivity, specific, positive predictive value (PPV) and negative predictive value (NPV) were calculated using standard methods. In keeping with prior studies in both ARFI and TE,[28,126,175] a difference of ≥2 F-score between the ARFI and Metavir F scores was deemed clinically significant and used to define discordance between ARFI LSM and biopsy. Deviation from biopsy was calculated as the difference in velocity (m/s) of the ARFI LSM from the histopathologic reference range. Hence a patient with an ARFI LSM of 1.50m/s but F1 disease on histopathology (i.e. equivalent to an ARFI cut-off ≤1.34m/s) would be defined as deviating 0.16m/s from biopsy.

The relationship between continuous variables was assessed using the Spearman's rank correlation. Multi-regression analyses were performed, with equal weighting applied to all predictor variables. The unpaired T test and Mann Whitney test were used for the comparison of groups of continuous variables, predominantly when following a parametric and non-parametric distribution respectively. The Kruskal-Wallis test was performed for assessing differences between three groups. The test employed is explicitly stated with each result, and the output from the statistical program and associated normality testing is provided as appendices. Normality was assessed using the D'Agostino-Pearson method and quantile-quantile plot (Q-Q plot). Comparison of categorical data was performed using the Fisher's exact test when comparing two groups, and Chi-square when comparing three groups. A p-value of <0.05 was used to define statistical significance throughout the study.

3.2 2D-SWE (Toshiba)

3.2.1 Ethics

The 2D-SWE study was approved by the Melbourne Health Research Ethics Committee as a low-risk project (HREC reference number: QA2014151). All participants provided verbal consent prior to study involvement, following explanation regarding the research and provision of a plain language statement (Appendix 1).

3.2.2 Patient cohort

Fifty-five patients with diffuse chronic liver disease of variable aetiology were prospectively enrolled in the study. Participants were consecutively recruited from the hospital's radiology department, having been clinically referred for the non-invasive assessment of liver fibrosis using existing elastography technology in clinical use in the institution (ARFI, Siemens). Patients were required to be over 18 years of age for study enrolment. There were no further inclusion criteria relating to patient demographics, CLD aetiology or disease severity.

Patient clinical information was obtained prospectively from hospital medical records at the time of 2D-SWE examination. Information collected included patient age at time of 2D-SWE, gender and CLD aetiology (as per physician assessment documented in the medical records). Height and weight measurements were recorded at the time of scanning for use in Body Mass Index calculations (BMI = weight (kg) / height (meters)^2). The presence and severity of hepatosteatosis was graded according to the level of echogenicity and beam attenuation observed on B-mode ultrasound, as discussed in Chapter 3.1.[30] Liver biopsy results were not available for the patient cohort and Toshiba 2D-SWE accuracy was therefore not assessed.

3.2.3 Toshiba 2D-SWE system

Patients completed the 2D-SWE measurements during the same session as ARFI and a targeted B-mode ultrasound examination of the liver. 2D-SWE measurements were

acquired with the Aplio 500 Platinum Series ultrasound machine (Toshiba Medical Systems Corporation, Tochigi, Japan); which is also a general purpose system designed for wide-ranging clinical ultrasound applications in addition to 2D-SWE (Figure 3.6). Measurements were acquired with the PVT-375BT probe (6-1.9MHz); a convex probe designed for abdominal use.

Figure 3.6: The Aplio 500 ultrasound machine from Toshiba Medical Corporation, used for the acquisition of all 2D-SWE measurements. (Image reproduced from the Toshiba website).

3.2.4 Operators

Measurements were acquired by a single operator, which included one of three experienced sonographers. All three operators received basic training in 2D-SWE and the Aplio 500 system prior to study commencement. The operators however had significant prior point SWE experience, each completing in excess of 100 ARFI scans previously.

3.2.5 Measurement acquisition

Patients were fasted for six hours prior to testing and measurements were acquired in the supine position. The scan was performed via an intercostal approach, with the right arm abducted to increase the intercostal window. Measurements were acquired with breathing held in light suspension. The ultrasound probe was positioned firmly against the skin with transmission gel, whilst avoiding any pressure which may compress the underlying liver.

Samples were taken in one region of the right hepatic lobe with a good acoustic window. The ultrasound plane was positioned perpendicular to the liver capsule to minimize ultrasound beam refraction. Regional shear wave propagation characteristics were initially qualitatively assessed using the continuous imaging mode. Once an acoustic window suitable for quantitative assessment was identified, a 'single shot' image of the liver was then obtained. This provides a static high-quality image of shear wave properties within an arc shaped section of liver measuring approximately 35mm in maximal width and 30mm in axial depth (Figure 3.7). The circular measurement region of interest (ROI) of fixed 10mm diameter was subsequently positioned in an area of liver suitable for quantitative assessment. This included areas of liver parenchyma away from vascular or biliary structures, which had uniform shear wave propagation characteristics on the Propagation Map and Speed Smart Map (described in Chapter 2.7). Areas ideal for quantitative assessment were indicated by parallel lines on the Propagation Map and relatively homogeneous color on the Speed Smart Map. Examples of recommended ROI positioning on the two display modes are demonstrated in Figure 3.8. Note was made to avoid areas of non-filling on the Speed Smart Map which indicate unreliable regional shear wave characteristics; similar to other 2D-SWE systems. On the occasion that a 'single shot' acquisition showed no areas suitable for quantitative assessment, the 'single shot' image was reacquired. Ten 2D-SWE measurements were obtained per patient, one per 'single shot' acquisition. Operators were blinded to all individual measurement velocities during the acquisition process via a screen shield.

The device automatically calculates the mean velocity (meters/second) and Young modulus (kPa) of shear waves within the chosen ROI for each measurement; the former recommended by the manufacturer for liver fibrosis quantification. The standard deviation (SD) of shear wave velocities obtained within the measurement ROI is also automatically generated and displayed for each 2D-SWE reading (Figure 3.7). The ratio of the ROI SD to the mean shear wave velocity of the corresponding measurement (ROI SD/Speed) was

calculated for all ten readings per patient. ROI SD/Speed was assessed as a potential metric for assessing individual measurement reliability, akin to IQR/Median being used to assess the reliability of measurement sets overall. The Skin-to-Liver Capsule Distance (SLD) was manually measured to three decimal points using the picture archiving and communication service (PACS) ruler function (Figure 3.7). The distance from the liver capsule to ROI center was also manually measured for each individual measurement.

Figure 3.7: Screenshot of a Toshiba 2D-SWE 'single shot' acquisition, which shows a static image of shear wave propagation characteristics within an arc shaped 'field of view'. A round 10mm measurement ROI is then positioned in an area of parenchyma deemed suitable for quantitative assessment. The shear wave velocity and standard deviation of shear wave velocities within the measurement region of interest (ROI SD) are automatically generated and displayed (bottom of the image). The Skin-to-Liver Capsule Distance (SLD) and Liver Capsule to ROI Depth were manually measured as indicated.

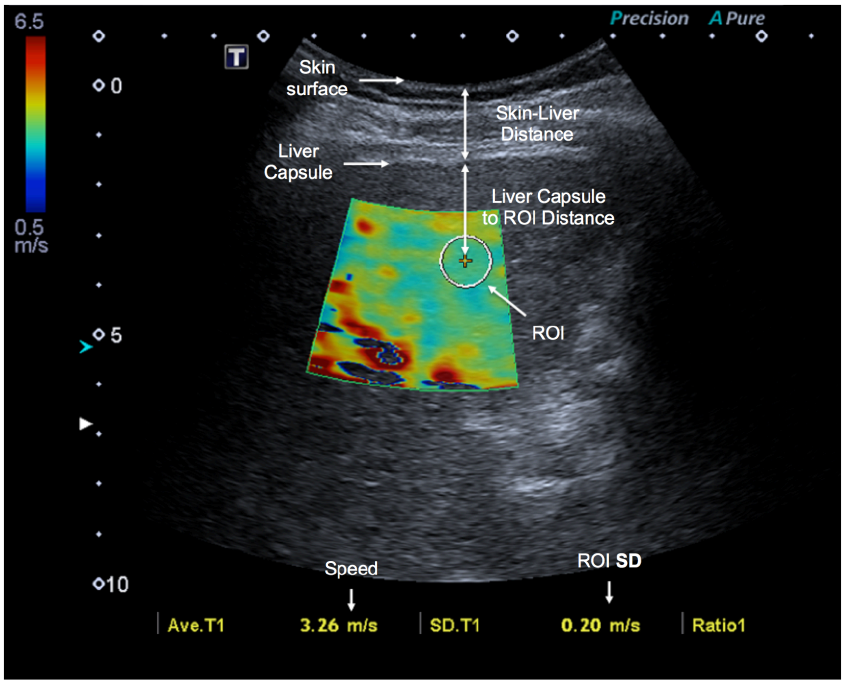
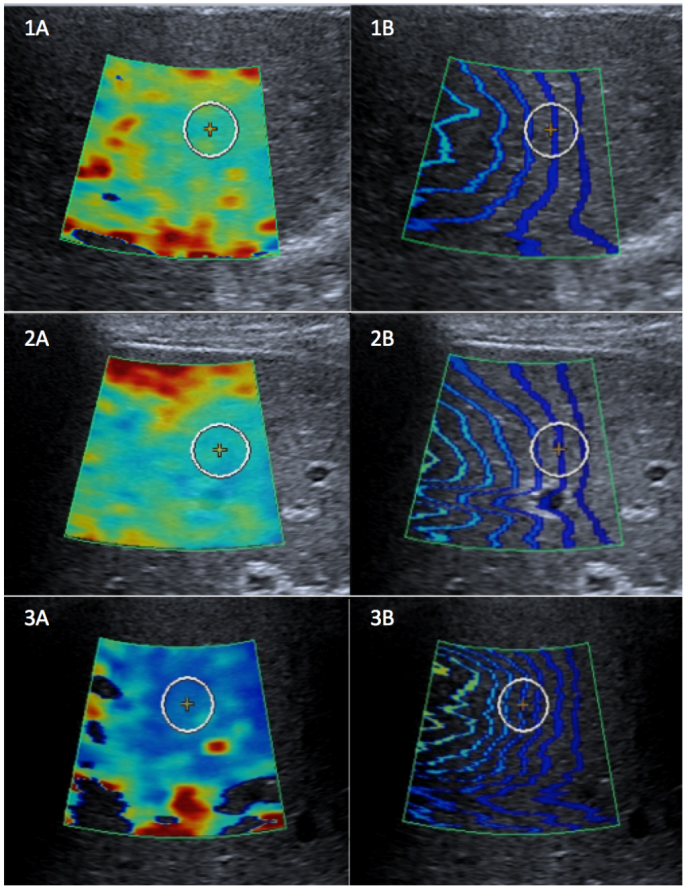


Figure 3.8: Speed Smart Maps (1A, 2A, 3A) show the distribution of shear wave velocities through a section of liver; red areas representing high velocities and blue/green areas low velocities. Propagation Maps (1B, 2B, 3B) show the arrival time contours of shear waves at different points in the liver. The region of interest (ROI) is positioned in an area with uniform shear wave characteristics; as indicated by homogeneous color on the Speed Smart Map and parallel contour lines on the Propagation Map.



3.2.6 Statistical analyses

Statistical analyses were completed using Graphpad prism 8 and Stata 12 statistical software. The statistical approaches employed are similar to those described in the ARFI methods chapter.

The interquartile range to median ratio (IQR/Median) was calculated for 2D-SWE measurement sets, as per standard methods in ARFI and TE. Cronbach's alpha was used to determine the internal consistency of readings and Bland Altman plot for the analysis of

optimal measurement number.[202,203] The closeness in approximation to the set's median of 10 samples achieved with differing numbers of acquired/analysed measurements was assessed. A pre-defined deviation threshold of <5% was used to denote a precise LSM, as this threshold has been used in the ARFI and SSI literature.[95,204] The distribution of the ten measurements acquired within each patient was assessed using skewness and kurtosis.

The relationship between patient factors and IQR/Median and ROI SD/Speed values were assessed using Spearman's rank correlation and the Kruskal-Wallis test (the latter used predominantly for the table data). The Mann Whitney and Fisher exact tests were used to compare groups of continuous and categorical data, respectively, with the test employed explicitly documented with each result Normality testing and statistical program outputs are included in Appendix 2.

The impact of factors affecting the reproducibility of individual measurements (i.e. Capsule to ROI Depth or ROI SD/Speed) were assessed by calculating the deviation of individual 2D-SWE measurements from the set's overall median velocity of 10 readings.

A p-value of <0.05 was used to denote statistical significance.

Intra-operator and inter-operator reliability were not assessed in the study. Liver histopathology was not available for this subset of the study to allow the evaluation of Toshiba 2D-SWE accuracy.

Chapter 4. ARFI Results

4.1 Patient Characteristics

4.1.1 Greater ARFI cohort

The overall ARFI cohort was heterogeneous in respects to patient demographics and chronic liver disease characteristics. The median age of the cohort was 51 years (range: 16 to 89 years), with 51% of patients being male. A large proportion of patients were overweight (BMI = 25 - 30kg/m^2, 32%) or obese (BMI >30kg/m^2, 26%).

A wide range of CLD aetiologies was encompassed within the greater ARFI cohort (Table 4.1), with multiple contributory pathologies present in a significant proportion of patients. The most common CLD aetiologies included NAFLD (27.0%), Hepatitis B (25.7%), Hepatitis C (21.3%) and alcohol (11.4%). As graded by B mode imaging, significant hepatosteatosis (i.e. moderate to severe in severity) was observed in 28.0% of patients. The breakdown of patient characteristics is listed in Table 4.1.

Table 4.1: Patient demographic and CLD characteristics of the overall ARFI patient cohort.

Patient Demographics	**No. of patients (% of cohort)**
Gender	
Male	481 (51.0%)
Female	463 (49.0%)
Age (years)	
<20	11 (1.2%)
20 – 39	222 (23.6%)
40 – 59	441 (46.7%)
60 – 79	248 (26.3%)
≥80	21 (2.2%)
BMI (kg/m^2)	
<20	13 (5.2%)
20-25	91 (36.7%)
25-30	80 (32.2%)
30-35	40 (16.1%)
>35	24 (9.7%)

CLD Aetiology (based on clinician assessment)	
Non-alcohol fatty liver disease (NAFLD)	255 (27.0%)
Hepatitis B	243 (25.7%)
Hepatitis C	201 (21.3%)
Alcohol	108 (11.4%)
Drug-induced liver injury	47 (5.0%)
Autoimmune Hepatitis	43 (4.6%)
Primary Biliary Cholangitis (PBC)	34 (3.6%)
Haemochromatosis	32 (3.4%)
Primary Sclerosing Cholangitis (PSC)	20 (2.1%)
Cardiac Cirrhosis	17 (1.8%)
Porphyria	10 (1.1%)
Wilsons' Disease	5 (0.5%)
Cryptogenic	22 (2.3%)
Other	31 (3.3%)
Hepatosteatosis (based on B-mode U/S)	
No	377 (52.7%)
Mild	137 (19.2%)
Moderate	119 (16.6%)
Severe	82 (11.4%)

4.1.2 Liver Biopsy sub-cohort

Fifty-five patients in the study cohort had a recent liver biopsy result which fulfilled the review criteria outlined in Chapter 3.1.7. The biopsy was performed a mean of 34.7 days from ARFI, however was completed on the day of imaging in 29 patients (i.e. 52.7% of the cohort). The median biopsy length was 20mm (range: 15 to 26mm) and included a median of 12 portal tracts (range: 6 to 24 tracts).

Reflecting the clinical indications for biopsy, there were some differences in the patient characteristics between the liver biopsy sub-cohort and the greater ARFI cohort (Table 4.2). In particular, the liver biopsy sub-cohort comprised a higher proportion of patients with Autoimmune Hepatitis (AIH), Drug Induced Liver Injury (DILI), Primary Biliary Cholangitis (PBC) and Primary Sclerosing Cholangitis (PSC). The liver biopsy sub-cohort also had a higher median ALT of 61 IU/L (q1-q3: 30 – 106 IU/L) than those in the greater ARFI cohort (median ALT = 38 IU/L, q1-q3: 24 – 69 IU/L).

Table 4.2: CLD aetiology amongst the liver biopsy sub-cohort.

CLD Aetiology (as per clinician assessment)	**No. of patients (% of sub-cohort)**
Non-alcoholic Fatty Liver Disease (NAFLD)	15 (27%)
Hepatitis B	8 (15%)
Autoimmune Hepatitis	8 (15%)
Hepatitis C	6 (11%)
Drug induced liver injury	6 (11%)
Primary Biliary Cholangitis	6 (11%)
Haemochromatosis	3 (5%)
Primary Sclerosing Cholangitis	3 (5%)
Alcohol	1 (2%)
Other	5 (9%)

Table 4.3: Breakdown of histopathology findings amongst the 55 patients in the liver biopsy sub-cohort.

	Number of patients (% of cohort)
Metavir F score	
- F0/1 (no / mild fibrosis)	29 (53%)
- F2 (significant fibrosis)	12 (22%)
- F3 (severe fibrosis)	6 (11%)
- F4 (cirrhosis)	8 (15%)
Metavir A score	
- A0 (no inflammation)	20 (36%)
- A1 (mild)	17 (30%)
- A2 (moderate)	14 (25%)
- A3 (severe)	4 (7%)
Brunt steatosis score (i.e. % of hepatocytes with macrosteatosis)	
- <5%	25 (45%)
- 5 – 33%	19 (35%)
- 33 – 66%	5 (9%)
- >66%	6 (11%)

There was a wide range in histopathologic findings amongst the liver biopsy sub-cohort (Table 4.3). Twenty-nine patients (53% of the cohort) had either no or mild fibrosis (i.e. Metavir F0/F1), whilst conversely eight patients (15%) had cirrhosis (i.e. Metavir F4). There was either moderate or severe necroinflammatory change (Metavir A2 / A3) in 18 patients (32% of the cohort). Steatotic change involved a median of 5% of hepatocytes on biopsy, however moderate or severe hepatosteatosis was observed in 11 patients (i.e. 20% of the cohort).

Inter-operator agreement between the fibrosis stages of the blinded pathologist and the original clinical report could not be reliably assessed due to differences in scoring systems employed. Namely, the biopsy specimens were regarded using the Metavir scoring system[73] whilst the original clinical report staged specimens according to Scheuer.[199]

4.1.3 Cirrhosis sub-cohort

One hundred and eighty-six patients (i.e. 19.7% of the overall cohort) had a clinical diagnosis of cirrhosis at the time of ARFI scanning on the basis of non-elastography parameters (outlined Chapter 3.1.4). Of these patients, eighty-five (45%) had clinical or imaging evidence of portal hypertension at the time of ARFI assessment.

After applying the review criteria outlined in Chapter 3.1.3, valid MELD scores were available for 66 patients and valid Child-Pugh Scores for 61 patients. The median MELD score was 8.5 (range = 6 – 20.3). The breakdown of the MELD and Child-Pugh Scores is listed in Table 4.4. As only four patients had Child-Pugh C cirrhosis, Grades B and C were combined together for all subsequent analyses.

Table 4.4: Breakdown of Child-Pugh and MELD scores within the sub-cohort of patients with cirrhosis.

	No. of patients (% of cohort)
Child-Pugh Score	
- 5	22 (36%)
- 6	18 (30%)
- 7	7 (11%)
- 8	7 (11%)
- 9	3 (5%)
- 10 / 11	4 (7%)
Child-Pugh Grade	
- A	40 (65%)
- B	17 (28%)
- C	4 (7%)
MELD Score	
- <8	32 (49%)
- 8 to 12	16 (24%)
- 12 to 16	13 (20%)
- >16	5 (7.6%)

The majority of patients had compensated cirrhosis, with a minority having complications resulting from portal hypertension or synthetic failure (Table 4.5). Only 12% of the sub-cohort had documented hepatic encephalopathy, whilst ascites was present in 20% of individuals. Amongst those with a recent gastroscopy (i.e. performed within 6 months of ARFI), 27 patients (i.e. 46% of cirrhotic patients) had evidence of oesophageal varices, of which 10 were medium or large in size.

Table 4.5: Frequency of cirrhotic complications within the cirrhotic sub-cohort of patients. The presence or absence of oesophageal varices only encompasses patients with a recent gastroscopy result (i.e. those performed within six months of ARFI).

	No. of patients (% of cohort)
Portal hypertension	
No	101 (54%)
Yes	85 (46%)
Oesophageal varices	
No	32 (54%)
Small	17 (29%)
Medium	7 (12%)
Large	3 (5%)
Gastric varices	
No	53 (90%)
Yes	6 (10%)
Gastrointestinal bleeding	
Yes	9 (5%)
Requiring blood transfusion	4 (2%)
Hepatic encephalopathy (Westhaven criteria)	
No	155 (88%)
Grade I / II	13 (7%)
Grade III / IV	8 (5%)
Ascites	
No	139 (80%)
Mild	24 (12%)
Moderate / Severe	10 (8%)
Hepatocellular Carcinoma	
No	169 (97%)
Yes	6 (3%)

4.2 ARFI scan characteristics

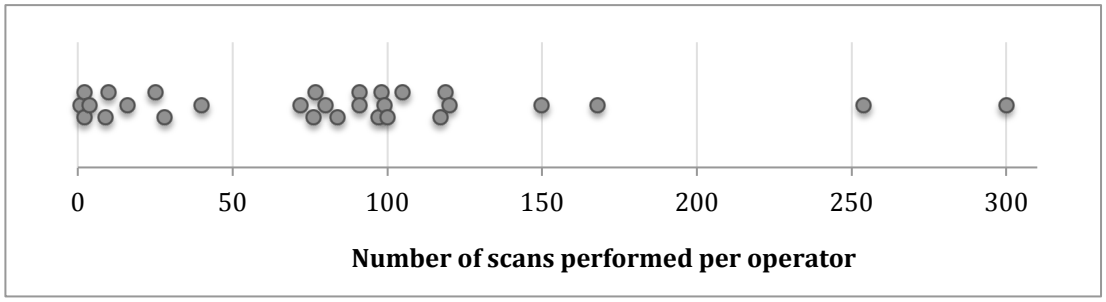
A total of 943 patients were scanned with ARFI during the study period encompassing 10/08/2012 to 09/12/2014. As described in Chapter 3.1, all patients were routinely scanned with either two or three operators, depending on their scan completion date. A small number of patients (n=10) had measurements taken by a single operator only due to logistical reasons on the day of scanning. A summary of the number of planned operators is listed in Table 4.6. Any additional (i.e. unplanned) ARFI measurements were not included in subsequent analyses.

Table 4.6: Number of pre-planned operators who completed ARFI measurements per patient.

Number of pre-planned ARFI measurements	Number of patients (% of cohort)
1 Operator	10 (1.1%)
2 Operators	642 (68.1%)
3 Operators	291 (30.8%)

There were a total of twenty-nine different operators over the study period, who had a broad range in ARFI experience (i.e. cumulative number of ARFI scans completed). The breakdown of individual operator experience is shown in Figure 4.1.

Figure 4.1: Total number of scans performed per operator by study conclusion.



Of the total 2,165 ARFI scans performed amongst the cohort's 943 patients, an ARFI velocity was able to be obtained in nearly all patients (i.e. 2,161, 99.8% of scans). The frequency of ARFI velocities observed amongst these 2,161 scans is shown in Figure 4.2. After applying the F score cut-offs based from the meta-analysis of Friedrich Rust *et al.,*[145] the majority of ARFI scans had elastography results which fell in the F0/F1 (50.5%) and F4 (28.8%) categories. The individual breakdown is shown in Table 4.7.

Figure 4.2: Frequency of ARFI velocities observed in the overall patient cohort.

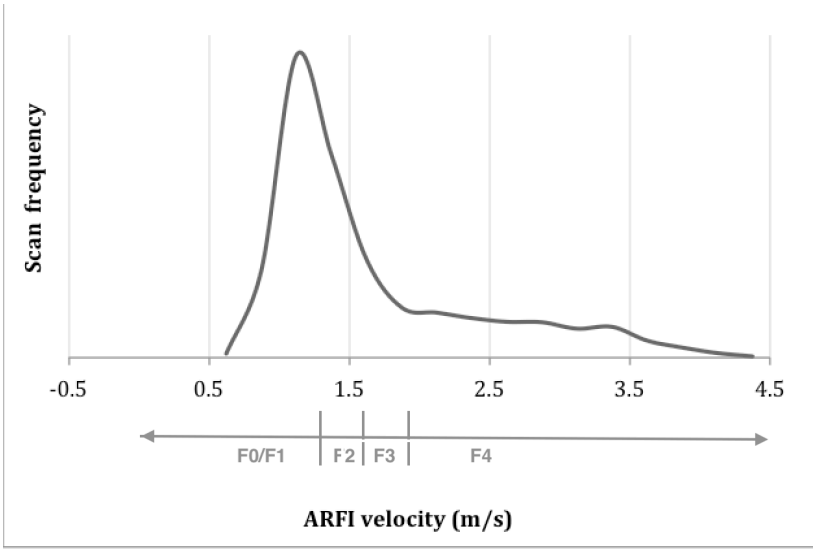


Table 4.7: Breakdown of ARFI results falling into each F score category.

F score (as per ARFI scan)	Number of scans (% of cohort)
F0/F1	1094 (50.5%)
F2	278 (12.9%)
F3	169 (7.8%)
F4	624 (28.8%)

4.3 ARFI performance

4.3.1 Accuracy

ARFI velocities showed only a moderate correlation with biopsy F scores (rho = 0.284, p = 0.003) and was concordant with histopathology in 71.0% of patients (95%CI: 61.8 – 78.8%). The obtained ARFI velocity deviated a mean of 0.39 m/s (95%CI: 0.29 – 0.50 m/s) from the reference range of the biopsy F score. The mean net deviation was + 0.26 m/s (95% CI: 0.14 – 0.38 m/s), indicating that the majority of inaccurate measurements deviated above the biopsy reference range (i.e. ARFI overestimating fibrosis levels). The AUROC, sensitivity, specificity, PPV and NPV of ARFI in discriminating liver fibrosis at the three different F score cut-offs is listed in Tables 4.8 and 4.9 below. Overall, ARFI showed a good sensitivity and NPV, but had relatively poor specificity and PPV; again reflecting the technology's propensity for false positive readings.

Table 4.8: AUROC of ARFI in discriminating the 3 different F score cut-offs.

F score cut-off	AUROC (95% CI)
F0/F1 vs. ≥F2	0.67 (95%CI: 0.56 – 0.77)
≤F2 vs. ≥ F3	0.76 (95%CI: 0.66 – 0.86)
≤F3 vs. F4	0.70 (95%CI: 0.57 – 0.83)

Using AUROC analysis, we calculated optimal cut-offs to be 1.64, 1.78 and 1.95m/s for the F1/F2, F2/3 and F3/4 cut-offs respectively (i.e. all higher than those drawn from the meta-analysis of Friedrich Rust *et al*). Applying these cut-offs resulted in a mild improvement in accuracy, however this was at the expense of sensitivity (Table 4.9).

To help assess the impact of referral bias (i.e. patients being referred for biopsy on the basis of unexpected ARFI results), accuracy was also assessed amongst only patients who completed the liver biopsy prior to or on the day of ARFI (n=34). These patients showed only mildly stronger correlation between ARFI and histopathology (rho = 0.370 vs. 0.284).

The AUROC at the three F score cut-offs also showed only mild improvement; having an AUROC = 0.70, 0.76 and 0.69 at detecting ≥F2, ≥F3 and F4 disease, respectively.

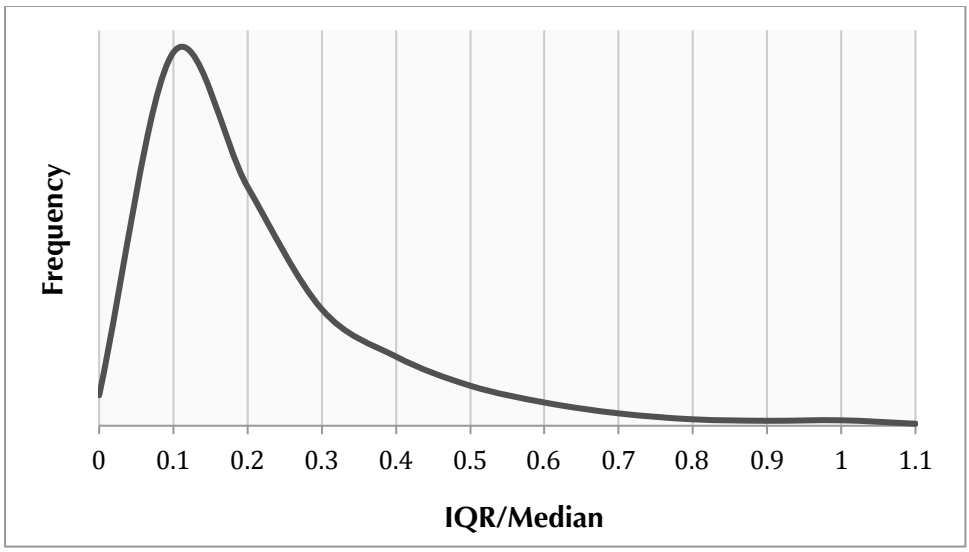
Table 4.9. Sensitivity, specificity, PPV and NPV of ARFI at discriminating between the three different F score cut-offs (biopsy Metavir F score being the reference).

F score cut-off	Accuracy (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)
Applying cut-offs from meta-analysis of Friedrich Rust *et al.*					
F1 vs. ≥F2 (1.34 m/s)	60.7% (52.8 - 68.2%)	80.4% (67.4 – 89.2%)	42.9% (20.8 – 55.9%)	56.2% (44.8 – 70.0%)	70.6% (53.7 – 83.3%)
≤F2 vs. ≥F3 (1.55 m/s)	67.3% (57.5 – 74.3%)	88.9% (71.1 – 97.0%)	60.0% (49.0 – 70.1%)	42.9% (30.8 – 55.9%)	94.1% (82.5 – 98.6%)
≤F3 vs. F4 (1.80 m/s)	68.2% (58.9 – 76.3%)	80.0% (54.1 – 93.7%)	66.3% (56.1 – 75.2%)	27.9% (16.6 – 42.8%)	95.3% (86.6 – 98.9%)
Applying optimized ARFI cut-offs (outlined below)					
F1 vs. ≥F2 (1.64 m/s)	67.5% (55.1 – 72.9%)	64.7% (51.0 – 76.4%)	64.3% (51.2 – 75.6%)	62.3% (48.8 – 74.1%)	66.7% (56.9 – 81.4%)
≤F2 vs. ≥F3 (1.78 m/s)	73.8% (64.7 – 81.3%)	77.8% (58.9 – 89.7%)	72.5% (61.8 – 81.1%)	48.8% (34.6 – 63.3%)	90.6% (80.7 – 96.0%)
≤F3 vs. F4 (1.95 m/s)	70.0% (60.8 – 78.0%)	73.3% (47.7 – 89.5%)	69.6% (59.5 – 78.1%)	28.2% (16.4 – 43.9%)	94.1% (85.4 – 99.1%)

4.3.2 Measurement variability

The IQR/Median ratio represents the amount of deviation between the 10 individual measurement velocities obtained within each measurement set. This value has been found to correspond to ARFI reliability, with lower accuracy observed amongst scans with an IQR/Median >0.3.[31,36,140] Of the 2,117 measurement sets, 329 (i.e. 15.5%) had IQR/Median parameters which 'failed' the IQR/Median reliability criteria. The median IQR/Median value for the cohort was 0.17 (q1-q3: 0.10 – 0.29), with the breakdown in values shown in Figure 4.3.

Figure 4.3: Breakdown of IQR/Median ratios of the patient cohort; the majority of patient's having an IQR/Median <0.30.



4.3.3 Inter-operator agreement

There was some deviation in the overall ARFI velocities obtained between the first two operators for each patient. The median difference in ARFI velocity between the first two operators was 0.17 m/s (q1-q3: 0.07 – 0.42m/s), whilst the median percentage deviation was 11.7% (q1-q3: 5.3 – 25.1%). The breakdown of percentage deviation within the patient cohort is outlined in Table 4.10. Defining operator concordance as the two operators having ARFI velocities within adjacent F score ranges, operator concordance was observed in 86.7% of patients (95%CI: 84.3 – 88.7%).

Inter-operator correlation co-efficient (ICC) was unable to be performed for the entire ARFI cohort, as patients were tested by a wide variety of operator pairs. Limited ICC analyses were instead performed amongst operator pairs who had regularly acquired measurements together (i.e. ≥20 patients). Amongst the four operator pairs who met this criteria, the median ICC was 0.82 (range = 0.54 – 0.94).

Table 4.10: Degree of deviation between operator pairs in the patient cohort. Significant deviation (i.e. >20%) between operator ARFI values was observed in a significant proportion (i.e. almost 30%) of patients.

% Deviation between operator ARFI scores	Number of patients (% of cohort)
<5%	223 (24.0%)
5 to 10%	192 (20.6%)
10 to 20%	228 (24.9%)
20 to 40%	169 (18.2%)
40 to 60%	68 (7.3%)
>60%	51 (5.5%)

4.4 Patient factors affecting ARFI performance

The impact of a number of patient factors on ARFI IQR/Median, inter-operator concordance and accuracy (i.e. correlation with biopsy) is summarised in Tables 4.11, 4.12, 4.13 and 4.14 and discussed in the relevant sections below.

Table 4.11: Impact of patient factors on IQR/Median and inter-operator reliability (i.e. percentage deviation between operators and inter-operator concordance) for the overall ARFI cohort. Increasing age, BMI, SLD, hepatosteatosis, ALT and a diagnosis of NAFLD were all associated with lower ARFI reliability (i.e. increased IQR/Median ratios and lower inter-operator agreement).

	IQR/Median (Median, q1, q3)	Median deviation between operator velocities (%)	Inter-operator concordance (%)
Age (years)			
- <40	0.13 (0.12, 0.21)	9.0% (3.7%, 18.3%)	89.1%
- 40 to 59	0.17 (0.11, 0.29)	11.3% (5.1%, 25.0%)	87.8%
- 60 to 79	0.21 (0.13, 0.36)	15.4% (7.5%, 30.5%)	83.3%
- ≥80	0.22 (0.14, 0.38)	14.3% (7.5%, 28.4%)	85.7%
Gender			
- Male	0.17 (0.10, 0.29)	11.5% (5.6%, 25.6%)	88.4%
- Female	0.17 (0.11, 0.29)	11.9% (4.9%, 25.0%)	85.4%
CLD Aetiology			
- NAFLD	0.20 (0.13, 0.36)	15.4% (6.6%, 33.3%)	82.6%
- Hepatitis B	0.14 (0.09, 0.24)	9.1% (4.2%, 19.2%)	91.2%
- Hepatitis C	0.17 (0.11, 0.26)	11.9% (6.2%, 24.8%)	86.5%
- Alcohol	0.19 (0.11, 0.30)	10.3% (6.2%, 20.7%)	93.3%
BMI (kg/m^2)			
- <25	0.12 (0.79, 0.18)	8.0% (3.5%, 12.6%)	95.2%
- 25 to 30	0.17 (0.11, 0.29)	14.1% (7.0%, 28.5%)	85.0%
- 30 to 35	0.24 (0.14, 0.37)	14.0% (7.4%, 32.0%)	75.0%
- >35	0.33 (0.20, 050)	19.9% (9.9%, 32.7%)	87.5%
SLD			
- < 2 cm	0.13 (0.10, 0.18)	8.3% (3.5%, 14.2%)	94.7%
- 2 to 2.5cm	0.20 (0.13, 0.31)	14.3% (6.5%, 31.1%)	83.4%
- >2.5cm	0.32 (0.23, 0.44)	21.9% (9.0%, 40.7%)	78.6%
Hepatosteatosis (on B-mode U/S)			
- No	0.15 (0.10, 0.25)	10.9% (4.5%, 20.9%)	87.7%
- Mild	0.18 (0.11, 0.36)	12.9% (6.1%, 27.6%)	86.9%
- Moderate	0.20 (0.13, 0.33)	12.6% (6.0%, 32.2%)	79.3%
- Severe	0.25 (0.15, 0.37)	19.6% (8.4%, 39.7%)	81.5%
ALT			
- < 56 u/L	0.16 (0.10, 0.29)	10.8% (5.1%, 22.5%)	87.2%
- 56 – 99 u/L	0.18 (0.12, 0.31)	14.8% (8.4%, 28.6%)	89.4%
- ≥ 100 u/L	0.19 (0.13, 0.33)	15.1% (8.5%, 36.3%)	86.4%

Table 4.12: Impact of patient factors on ARFI accuracy (i.e. concordance with histopathology F score and deviation from the biopsy reference range) amongst the liver biopsy sub-cohort. Increasing age, BMI, SLD and necroinflammatory activity were all associated with lower ARFI accuracy.

	Patient number (%)	% Concordant with biopsy F score	Mean deviation from the biopsy reference range (m/s)
Age (years)			
- < 40	16 (29%)	83.9%	0.27 m/s
- 40 to 59	25 (45%)	58.7%	0.40 m/s
- ≥ 60	14 (25%)	76.7%	0.44 m/s
Gender			
- Male	29 (53%)	66.7%	0.45m/s
- Female	26 (47%)	76.0%	0.28 m/s
CLD Aetiology			
- NAFLD	15 (27%)	64.5%	0.45 m/s
- Hepatitis B	8 (15%)	50.0%	0.43 m/s
- Hepatitis C	6 (11%)	71.4%	0.37 m/s
- Autoimmune Hepatitis	8 (15%)	66.7%	0.61 m/s
- Primary Biliary Cholangitis	6 (11%)	100.%	0.01 m/s
- Drug induced liver injury	6 (11%)	75.0%	0.75 m/s
SLD			
- <2cm	18 (32%)	81.8%	0.21 m/s
- 2 to 2.5cm	25 (45%)	89.4%	0.34 m/s
- >2.5cm	12 (22%)	40.0%	0.94 m/s
Hepatosteatosis (Brunt Score)			
- Minimal	25 (45%)	70.2%	0.46 m/s
- Mild	19 (35%)	48.3%	0.19 m/s
- Moderate / Severe	11 (20%)	63.6%	0.50 m/s
Necroinflammatory activity (Metavir A Grade)			
- A0 (none)	20 (36%)	68.4%	0.33 m/s
- A1 (mild)	17 (30%)	72.7%	0.25 m/s
- A2 / A3 (moderate / severe)	18 (32%)	72.2%	0.54 m/s

Table 4.13: Multi-regression analysis looking at the relative impact of patient factors on both IQR/Median and inter-operator agreement (i.e. percentage deviation between operators). Of the associations identified in Table 4.11, SLD appeared to be the primary independent factor determining IQR/Median and inter-operator agreement.

Patient Factor	Correlation with IQR/Median	Correlation with % deviation between operators
Age	$R^2 = 0.113$	$R^2 = 0.085$
Gender	$R^2 = -0.022$	$R^2 = -0.047$
BMI	$R^2 = -0.044$	$R^2 = -0.051$
SLD	$R^2 = 0.440$	$R^2 = 0.305$
Hepatosteatosis	$R^2 = -0.004$	$R^2 = 0.010$
ALT	$R^2 = 0.043$	$R^2 = 0.037$
Diagnosis of NAFLD	$R^2 = 0.004$	$R^2 = 0.017$

Table 4.14: Multi-regression analysis looking at the relative impact of patient factors on ARFI accuracy (i.e. deviation from the histopathology reference range). This too showed SLD to be predominant independent factor influencing ARFI accuracy, with a weaker association seen for necroinflammatory change (i.e. Metavir A score).

Patient Factor	Correlation with the gross deviation of ARFI from the biopsy reference range (m/s)
Age	$R^2 = -0.075$
Gender	$R^2 = 0.041$
SLD	$R^2 = 0.543$
Steatosis (% hepatocytes)	$R^2 = -0.010$
Necroinflammatory change (i.e. A score)	$R^2 = 0.167$

4.4.1 Patient demographics

ARFI reliability appeared to reduce with increasing patient age, with a statistically significant correlation noted between age and IQR/Median (rho = 0.246, p=0.001) and between age and the amount ARFI deviated from the biopsy reference range (rho = 0.188, p=0.001). An independent correlation between age and ARFI reliability was not however demonstrated in multi-regression analyses (Tables 4.13 and 4.14). The apparent association is therefore likely attributable to the change in body habitus with increasing age; body habitus being the primary factor determining ARFI reliability in multi-regression analyses (as discussed in subsequent sections). Age showed a statistically significant positive correlation with both BMI (rho = 0.209, p=0.001) and SLD (rho=0.178, p<0.001).

A diagnosis of non-alcoholic fatty liver disease (NAFLD) was also associated with higher IQR/Median ratios (median IQR/Median = 0.206 vs. 0.154, Mann Whitney, p<0.0001) than in patients with non-NAFLD diagnoses. NAFLD patients also had lower inter-operator agreement, with higher percentage deviation between operators (median deviation = 15.4% vs. 10.7%, Mann Whitney, p<0.0001) and lower rates of inter-operator concordance (82.6% vs. 89.4%, Fisher's exact, p=0.007). An independent association was not, however, demonstrated in multivariate analyses (Table 4.13 and 4.14), and the relationship between NAFLD and ARFI reliability appeared to be similarly attributable to body habitus. As would be expected, a diagnosis of NAFLD was associated with a significantly higher median BMI (29.8 vs. 24.9, Mann Whitney, p<0.0001) and higher median SLD (2.42 vs. 2.00cm, Mann Whitney, p<0.0001) than those with non-NAFLD diagnoses.

4.4.2 Necroinflammatory change

The presence of necroinflammatory change on biopsy (i.e. Metavir A score) was associated with lower ARFI accuracy (i.e. AUROC) in discriminating between the three different Metavir F score cut-offs (Table 4.15). The Metavir A score also showed a positive and independent correlation with the amount ARFI deviated from the biopsy Metavir F score reference range (R^2=0.167, Table 4.14). A statistically significant difference in ARFI deviation from the biopsy reference range was not however demonstrated between patients with moderate or severe inflammation (i.e. A2/A3) compared to those with no/mild (A0/A1) inflammation (mean deviation = 0.539 vs. 0.326m/s, unpaired t test, p=0.08).

Table 4.15: Accuracy of ARFI at discriminating between the three different histopathological Metavir F scores cut-offs amongst patients with no / minimal necroinflammatory change (A0/A1, left column) and those with moderate or severe inflammation (A2/A3, right column). Patients with lower A scores overall showed higher accuracy.

F score cut-off (ARFI)	AUROC of patients with A0/A1 on histopath (95% CI)	AUROC of patients with A2/A3 on histopath (95% CI)
F0/F1 vs. ≥F2	0.58 (95%CI: 0.43 – 0.74)	0.53 (95%CI: 0.21 – 0.858)
≤F2 vs. ≥ F3	0.80 (95%CI: 0.69 – 0.91)	0.62 (95%CI: 0.42 – 0.82)
≤F3 vs. F4	0.86 (95%CI: 0.77 – 0.94)	0.50 (95%CI: 0.27 – 0.73)

Table 4.16: Multi-regression analysis looking at the correlation between patient factors and absolute ARFI velocity, after controlling for histopathological Metavir F score. SLD and to a lesser extent necroinflammatory change showed a positive correlation with ARFI velocity (i.e. independent of underlying liver fibrosis); indicative of a propensity towards increased liver stiffness / falsely positive ARFI velocities. Conversely, steatosis was independently associated with lower ARFI velocities (i.e. reduced liver stiffness).

Patient Factors	Correlation with absolute ARFI velocity (controlled for Metavir F score)
Age	R^2 = -0.030
Gender	R^2 = -0.043
SLD	R^2 = 0.475
Steatosis (% hepatocytes)	R^2 = -0.147
Necroinflammatory change (i.e. A score)	R^2 = 0.242 .

Multi-regression analyses were also performed to assess whether any patient factors impacted on absolute ARFI velocity, after controlling for the underlying histopathologic fibrosis level (i.e. Metavir F score, Table 4.16). This demonstrated a positive correlation between necroinflammatory change (i.e. A score) and absolute ARFI velocity, which was independent of underlying liver fibrosis (R^2 = 0.242); indicating a propensity towards higher liver stiffness values with increasing hepatic inflammation.

4.4.3 Hepatosteatosis

Hepatosteatosis was also associated with reduced ARFI reliability in univariate analyses. A statistically significant difference in IQR/Median ratios was observed between patients with no, mild or. moderate/severe steatosis; the median IQR/Median ratio being 0.134 vs. 0.189 vs. 0.231 amongst the three groups respectively (Kruskal-Wallis test, p<0.0001). Inter-operator deviation also increased in parallel with hepatosteatosis severity, with the median percentage deviation between operator ARFI velocities being 10.0%, 12.9% and 15.7% in patients with no, mild or moderate/severe steatosis respectively (Kruskal-Wallis, p<0.0001). This was also reflected in differences in inter-operator concordance rates, which were 88.8%, 86.9% and 90.8% for the three group, respectively (Chi-square test, p=0.031). These associations were not however shown to be independent in multi-regression analyses (Table 4.13), and are therefore again likely explained by the relationship between hepatosteatosis and body habitus (correlation between hepatosteatosis percentage and SLD: rho = 0.365, p=0.002). Furthermore, no independent association was observed between hepatosteatosis (i.e. % hepatocytes involved) and ARFI deviation from the biopsy reference range (rho=0.012, p=0.90, Table 4.12).

After controlling for other factors (including Metavir F score), hepatosteatosis showed a mild negative correlation with absolute ARFI velocity (R^2 = -0.147, Table 4.16). This suggests that the presence of hepatosteatosis results in a slight reduction in liver stiffness, independent of underlying liver fibrosis severity.

4.4.4 Body mass index

Body mass index (BMI) showed a moderately strong correlation with IQR/Median ratio (rho = 0.42, 95%CI: 0.34 – 0.49). The median IQR/Median ratio amongst obese patients (BMI >30 kg/m^2, IQR/Median = 0.300) was considerably higher than those in overweight (BMI: 25 – 30 kg/m^2, IQR/Median = 0.195) or normal / underweight (BMI<25 kg/m^2, IQR/Median = 0.127) categories (Kruskal-Wallis, p<0.0001). Inter-operator agreement was similarly impacted by body habitus, with BMI showing a statistically significant positive correlation with the percentage deviation observed between operator ARFI velocities (rho = 0.221, 95%CI: 0.136 – 0.303). Inter-operator concordance was observed in 95.2%, 85.0% and 81.5% of patients with a BMI ≤25, 25-30 and >30 kg/m^2 respectively (Chi Square, p=0.014). The relationship between BMI and ARFI accuracy could not be reliably assessed due to the small number of patients (n=5) in the biopsy sub-cohort with a contemporaneous BMI measurement recorded.

4.4.5 Skin-to-Liver Capsule Distance (SLD)

Skin-to-liver capsule distance (SLD) refers to the depth of abdominal wall tissue traversed by the ultrasound beam for a given ARFI measurement. This is measured from the skin surface to liver capsule and is primarily determined by the amount of subcutaneous tissue present in a patient. We found SLD to be the key determinate of ARFI performance, and consequently more detailed analysis of SLD as an ARFI parameter is included below.

There was a wide range in overall SLD values observed between patients (0.96 – 5.50cm), with the median SLD being 2.10cm. The breakdown of SLD values of patients in the cohort is listed in Table 4.17. SLD showed a very strong correlation with BMI (rho=0.799, p<0.0001, Figure 4.4), with an SLD of 2.5cm being approximately equivalent to a BMI of 30 kg/m^2.

Figure 4.4: SLD showed a strong linear correlation with BMI, demonstrating a rho of 0.799 (p<0.0001).

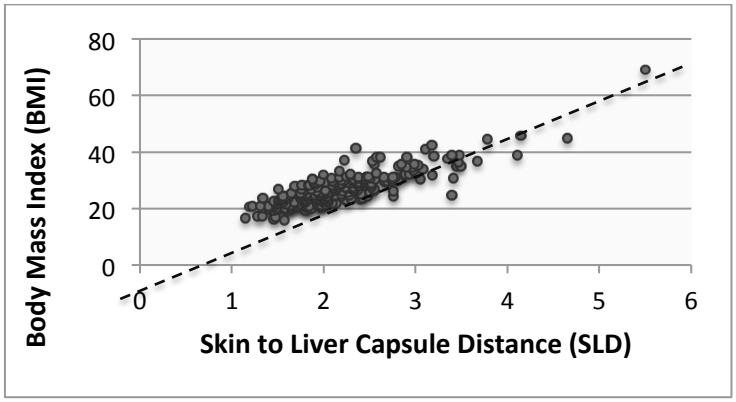


Table 4.17: Frequency of skin-to-liver capsule depths (SLDs) in the patient cohort, with their equivalent BMI.

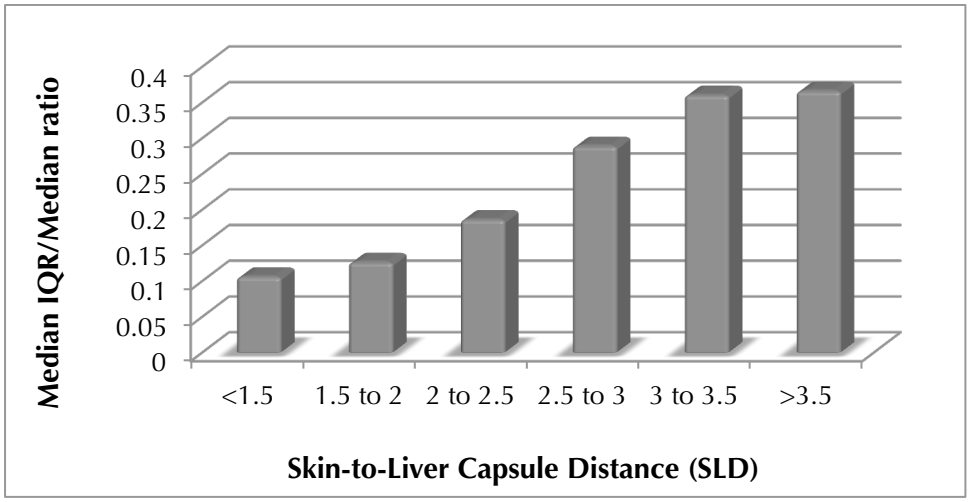
SLD	Median BMI	Number of patients (% of cohort)
<2cm	22.86	399 (42.3%)
2 to 2.5cm	27.61	331 (35.1%)
2.5 to 3cm	31.14	134 (14.2%)
3 to 3.5cm	32.21	57 (6.1%)
>3.5cm	44.76	22 (2.3%)

There was a low spread in SLD values between the ten individual ARFI measurements obtained within each individual patient. The median spread (i.e. median IQR) of SLD values per measurement set was 0.095 cm.

A low variation in overall SLD values was also observed between operators for each patient. There was a median 0.12cm (q1-q3: 0.05–0.22cm) difference between the SLD values obtained by the first two operators for each patient, which equated to a median percentage difference of 5.99% (q1–q3: 2.71-10.57%). This suggests any difference in operator technique has limited impact on SLD values.

SLD showed a significant relationship with measurement dispersion, with longer SLDs being associated higher IQR/Median ratios (rho=0.45, p<0.0001, Table 4.11, Figure 4.5). This correlation was slightly stronger than seen between BMI and IQR/Median (rho=0.42, p<0.0001), and SLD was found to be the primary determinate of IQR/Median in multi-regression analysis (Table 4.13). The number of scans deemed 'unreliable' by established IQR/Median criteria[140] was 3.72%, 14.69%, and 40.33% amongst patients with an SLD <2cm, 2-2.5cm and >2.5cm respectively (Chi-square, p<0.0001).

Figure 4.5: Relationship between SLD (cm) and IQR/Median ratio; showing increasing IQR/Median ratios with increasing SLD.



A similar relationship was observed between SLD and inter-operator concordance, with concordance rates being 94.7%, 83.4% and 78.6% amongst patients with an overall SLD of <2cm, 2-2.5cm and >2.5cm respectively (Chi-square, p<0.0001). The percentage deviation between the first two operator's ARFI velocities showed a similarly strong correlation with SLD (rho=0.347, p<0.001) than with BMI (rho=0.351, p<0.001). In multi-regression analysis however, SLD was found to be the predominant factor influencing inter-operator deviation (Table 4.13).

ARFI showed a markedly stronger correlation with biopsy Metavir F scores amongst patients with an SLD ≤2.5cm (rho=0.493, p<0.0001) than those with an SLD>2.5cm (rho=-0.242, p=0.54). A higher proportion of patients with an SLD≤2.5cm had ARFI F scores that were concordant with biopsy (80.5% vs. 40.0%, Fisher Exact, p<0.001, Table

4.12). On an individual measurement level, readings with longer SLDs showed greater deviation from the equivalent biopsy reference F-score range (rho= 0.376, Table 4.18). Patients with an SLD>2.5cm showed overall net deviation above the reference Metavir F-score range, indicating 'false positive' results (Table 4.18 and Figure 4.6). This propensity towards falsely elevated fibrosis estimates with increasing SLD values was further supported by multi-regression analysis (Table 4.15). This showed SLD to have a moderate positive correlation with ARFI velocity (SLD R^2 = 0.475) after controlling for underlying fibrosis levels (i.e. Metavir F score). This indicates that increasing SLD is associated with higher ARFI velocities, independent of underlying liver fibrosis severity.

Table 4.18: Mean deviation of individual ARFI measurements from the histopathology F score reference range, according to SLD. Measurements showed exponentially greater gross deviation from the histopathology F score range with increasing SLDs. These measurements showed net deviation above the biopsy reference range; indicating a propensity to 'false positive' results amongst measurements with a higher SLD.

SLD (cm)	Gross deviation from biopsy F score (Mean +/- SD)	Net deviation from biopsy F score (Mean +/- SD)
<2	0.207 m/s (+/- 0.361)	+ 0.164 m/s (+/- 0.383)
2 to 2.5	0.340 m/s (+/- 0.501)	+ 0.242 m/s (+/- 0.555)
2.5 to 3	0.758 m/s (+/- 0.834)	+ 0.681 m/s (+/- 0.898)
>3	1.520 m/s (+/- 0.931)	+ 1.520 m/s (+/- 0.931)

Given the relatively stronger association between SLD and ARFI performance than was observed with BMI, further analyses into the impact of overall tissue depth (i.e. skin surface to measurement ROI) on ARFI performance were performed. The total ARFI depth showed significant positive correlations with IQR/Median (rho = 0.398, p<0.0001), a positive correlation with the percentage deviation between operators (rho = 0.250, p<0.0001) and the deviation of ARFI velocities from the biopsy F score reference range (rho = 0.234, p=0.02). In multi-variate analyses however (Table 4.19), these associations were not independent of, and are therefore likely attributable to SLD.

Figure 4.6: Deviation of individual ARFI measurements from the histopathology reference range, as a function of SLD. Points which lie on the X-axis reflect accurate ARFI measurements, whilst those above or below represent falsely elevated or low ARFI velocities, respectively. A greater proportion of 'false positive' ARFI results was observed with increasing SLD.

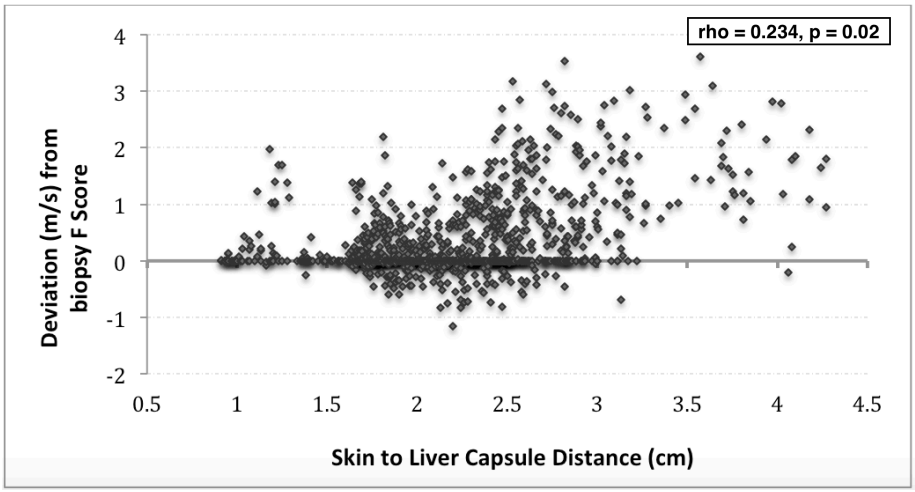


Table 4.19: Multi-regression analysis looking at the relative impact of SLD and total measurement depth on IQR/Median, inter-operator agreement (i.e. percentage deviation between operators) and the deviation of ARFI velocity from biopsy F score. Total measurement depth did not show a relationship with ARFI performance which was independent of SLD.

	Correlation with IQR/Median	**Correlation with % deviation between operators**	**Correlation with the gross deviation of ARFI from the biopsy reference range (m/s)**
SLD	$R^2 = 0.398$	$R^2 = 0.361$	$R^2 = 0.555$
Total depth	$R^2 = 0.076$	$R^2 = -0.064$	$R^2 = -0.026$

4.5 Scan factors affecting ARFI performance

4.5.1 Measurement Depth

The measurement ROI can be positioned at different depths within the liver for ARFI elastography assessment. The majority (98.9%) of measurements were taken over >1.5cm from the liver capsule, as per consensus guidelines.[140]

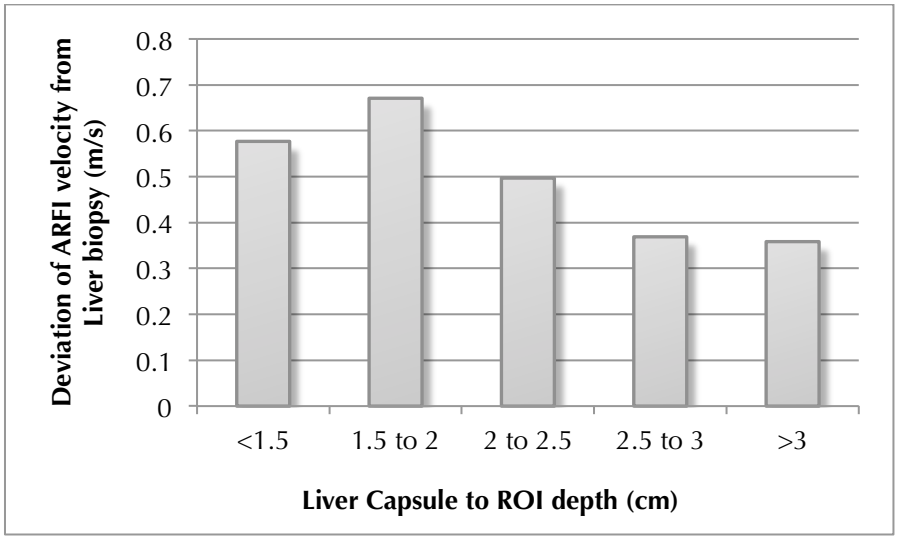
We found measurements performed within 2cm of the liver capsule (including those taken between 1.5cm and 2cm) showed greater deviation from the set's overall median value than those taken more deeply in the liver (Table 4.20). Measurements taken <2cm from the liver capsule deviated a mean of 0.31 m/s from the set's median velocity compared to 0.225m/s for measurements taken at depths >2cm (unpaired t test, $p<0.001$). This equated to a mean percentage deviation from the set's overall median velocity of 15.2% vs. 13.0% respectively (unpaired t test, $p<0.001$).

Table 4.20: Impact of measurement depth within the liver (i.e. liver capsule to ROI depth) on internal measurement consistency (i.e. deviation of the measurement from the set's overall median). Measurements taken within 2cm of the liver capsule showed reduced reliability.

Liver Capsule to ROI depth (cm)	Number of measurements (%)	Mean net deviation from set's median (m/s)	Mean gross deviation from set's median (m/s)	Mean gross % deviation from set's median
≤1.5	206 (1.1%)	0.069	0.368	15.7%
1.5 to 2	1,648 (28.1%)	0.046	0.305	15.2%
2 to 2.5	5,280 (33.9%)	0.030	0.243	13.6%
2.5 to 3.0	6,364 (19.0%)	0.023	0.218	12.9%
3.0 to 3.5	3,577 (6.2%)	0.010	0.207	12.1%
3.5 to 4.0	1,166 (6.2%)	0.002	0.229	13.6%
4.0 to 4.5	339 (1.8%)	- 0.018	0.230	14.0%
4.5 to 5.0	93 (0.5%)	- 0.060	0.264	14.3%
>5.0	130 (0.7%)	- 0.046	0.241	14.7%

Measurements taken within 2cm of the liver capsule also showed a poorer correlation with liver biopsy results (Figure 4.7). The mean deviation from the liver histology reference F score range was 0.655 vs 0.409 m/s in measurements taken <2cm and ≥2cm from the liver capsule respectively (unpaired t test, p<0.0001). The correlation between ARFI velocity and Metavir F scores was also stronger amongst measurements taken ≥2cm from the liver capsule (rho = 0.293 vs 0.051 respectively).

Figure 4.7: Mean deviation of individual readings from the histopathology reference score, according to measurement depth within the liver. Measurements taken <2cm from the liver capsule showed greater deviation from the biopsy reference range.



4.5.2 Individual operator experience

Individual operator experience (i.e. number of ARFI scans performed) showed only a weak negative correlation with IQR/Median ratio (rho = -0.047, p=0.03); indicating slightly higher IQR/Median ratios amongst inexperienced operators. A similar finding was observed with inter-operator reliability, with a weak negative correlation observed between individual operator experience and the percentage deviation between operators (rho = -0.074, p=0.002). The breakdown of these results is shown in Table 4.21.

Individual operator experience was also associated with lower ARFI accuracy (Table 4.22). Scans performed by operators with ≤25 scans experience showed significantly greater

deviation from the biopsy reference range than those performed by more experienced operators (mean deviation = 0.588 vs 0.301 m/s, unpaired t test, p=0.022). A lower proportion of ARFI scores also appeared to be concordant with histopathology amongst operators with ≤25 scans experience (72.0% vs. 51.6% respectively, Fisher's exact, p=0.07).

Table 4.21: IQR/Median ratio and inter-operator concordance rates according to individual operator experience (i.e. cumulative ARFI scans performed); demonstrating only a weak association between operator experience and ARFI reliability.

Number of scans performed per operator	IQR / Median (median, q1, q3)	% Of scans with concordant operators	% Deviation between operators LSMs (median, q1, q3)
11 to 25	0.170 (0.114 – 0.308)	84.2%	14.3% (6.3, 29.0%)
26 to 50	0.162 (0.108 – 0.288)	85.8%	12.3% (5.9, 25.7%)
51 to 75	0.186 (0.114 – 0.308)	87.5%	10.7% (5.6, 24.7%)
76 to 100	0.154 (0.106 – 0.289)	90.8%	10.7% (4.6, 19.2%)
101 to 150	0.142 (0.088 – 0.254)	86.6%	9.9% (4.9, 20.3%)
151 to 200	0.196 (0.112 – 0.312)	82.2%	12.8% (6.4, 26.9%)
>200	0.168 (0.096 – 0.287)	89.5%	9.9% (4.7, 27.3%)

Table 4.22: Correlation between ARFI and Metavir F score according to individual operator experience levels, showing greater deviation from the histopathologic reference range amongst operators with ≤25 scans experience.

Number of scans performed	% Of scans with ARFI velocities concordant with Biopsy	Mean deviation from biopsy (m/s)
≤25	51.6%	0.516 m/s
26 to 50	80.9%	0.253 m/s
51 to 100	62.5%	0.331 m/s
>100	77.3%	0.304 m/s

Sub-analyses were also performed amongst the most experienced operators (i.e. those who ultimately completed over 100 ARFI scans), to ensure the above findings were not solely attributable to the inclusion of transient, less proficient ARFI operators (e.g. visiting radiology / sonographer trainees) in the composite analyses. Amongst experienced operators (n=9), ARFI reliability still appeared lower amongst the first 25 scans performed. A negative correlation was again seen between individual operator experience and IQR/Median ratio (rho = - 0.10, p=0.002), and a weak negative correlation with inter-operator percentage deviation (rho = - 0.10, p=0.004). ARFI accuracy was also lower amongst the first 25 scans performed (mean deviation from biopsy = 0.575 vs. 0.260m/s, unpaired t test, p=0.024). The results breakdown is listed in Table 4.23.

Table 4.23: Relationship between individual operator experience and ARFI reliability (i.e. IQR/Median and inter-operator concordance) amongst experienced operators only (i.e those who ultimately completed over 100 ARFI scans).

Scans performed per operator	IQR / Median (median, q1, q3)	% Of scans with concordant operators	% Deviation between operators LSMs (median, q1, q3)
≤25	0.177 (0.118, 0.312)	83.5%	15.1% (6.5, 29.7%)
26 to 50	0.153 (0.106, 0.291)	87.3%	11.9% (5.3, 20.8%)
51 to 75	0.178 (0.117, 0.297)	87.6%	10.4% (5.9, 25.2%)
76 to 100	0.153 (0.106, 0.283)	91.0%	10.8% (4.8, 20.5%)

4.5.3 Institutional experience

There was a weak negative correlation between the cumulative number of patients scanned in the institution and the IQR/Median ratio (rho = -0.087, p<0.001). IQR/Median ratios were slightly higher amongst the first 150 patients scanned, however appeared to plateau thereafter (Table 4.24). The median IQR/Median ratio of the first 150 patients tested was 0.216 vs. 0.179 for subsequent scans (Mann Whitney, p=0.034).

Inter-operator concordance also appeared slightly lower during the early scans performed in the institution (Table 4.25). The percentage deviation between operators was slightly lower amongst the first 150 patients scanned (16.2% vs 11.0%, Mann Whitney test, p=0.0013). The difference between inter-operator concordance amongst the first 150 patients scanned compared to subsequent scans was not statistically significant (inter-operator concordance = 83.2% vs. 87.7%, Fisher's exact, p=0.14).

Table 4.24: Institutional experience (i.e. cumulative number of patients scanned) and IQR/Median ratios. There was slightly higher IQR/Median ratios and slightly higher failure of IQR/Median criteria amongst the first 150 patients scanned in the institution.

Cumulative patients scanned	**IQR/Median Median (Q1:Q3)**	**% Of scans passing 'IQR/Median≤0.30' criteria**
<50	0.214 (0.147, 0.414)	83.7% (75.7 – 90.3%)
50 to 99	0.212 (0.132, 0.367)	80.1% (71.9 - 87.4%)
100 to 149	0.220 (0.144, 0.341)	80.0% (71.0 – 86.7%)
150 to 249	0.184 (0.117, 0.311)	81.3% (75.3 – 86.1%)
250 to 499	0.167 (0.104, 0.287)	85.8% (82.0 – 88.9%)
500 to 749	0.192 (0.125, 0.311)	83.3% (79.7 – 86.3%)
≥750	0.171 (0.118, 0.296)	85.2% (82.1 – 88.7%)

Table 4.25. Impact of institutional experience on inter-operator agreement. Slightly greater inter-operator deviation and slightly lower inter-operator concordance was observed amongst the first 150 patients scanned in the institution.

Cumulative patients scanned	**% Discordance between Operators. Median (Q1,Q3)**	**% Scans with concordant operators (95% CI)**
<50	19.1% (8.5%, 31.6%)	80.0% (67.7 – 88.5%)
50 to 99	13.7% (7.0%, 26.1%)	85.8% (73.0 – 93.2%)
100 to 149	16.3% (8.1%, 29.0%)	82.0% (80.4 – 91.3%)
150 to 249	11.9% (3.9%, 25.4%)	89.1% (81.4 – 94.0%)
250 to 499	10.7% (4.6%, 21.5%)	90.0% (85.7 – 93.2%)
500 to 749	11.8% (5.8%, 27.2%)	84.3% (79.2 – 88.3%)
≥750	11.3% (5.6%, 25.1%)	88.4% (84.1 – 92.5%)

Amongst the small number of patients with a recent histopathologic correlate available, accuracy also appeared lower amongst the first 150 patients scanned in the institution (concordance with biopsy being 52.8% vs. 78.6% respectively, Fisher's test, p=0.15).

4.6 Quality Assessment Metrics

4.6.1 IQR / Median

ARFI measurements 'passing' established IQR/Median criteria (i.e. those with an IQR/Median ≤0.30) showed higher accuracy than those 'failing' these criteria. The correlation between ARFI velocity and histopathology was rho = 0.341 (p=0.002) vs. rho = -0.238 (p=0.3) amongst patients passing vs. failing the criteria respectively. Similarly, a higher proportion of scans that passed IQR/Median criteria were concordant with liver histopathology (77.9% vs. 42.9%, p=0.003). IQR/Median criteria had a sensitivity of 38.7% (95%CI: 23.7 – 56.2%) and a specificity of 88.2% (95%CI: 78.8 – 93.9%) in identifying clinically significant (i.e. ≥2 F score) discordance between ARFI and liver histology. The relative accuracy / sensitivity / specificity / PPV / NPV of patients passing the IQR/Median criteria at the three different F score cut-offs is listed in Table 4.26 below.

4.6.2 Skin-to-Liver Capsule Distance (SLD)

SLD was also assessed as a possible reliability metric, given its strong relationship with ARFI performance. Whilst ARFI performance appeared proportional to SLD across the full range of SLD values, ARFI reliability appeared to start deteriorating once SLD exceeded 2.5cm (Tables 4.11, 4.12 and 4.18). Hence a SLD ≤2.5cm was trialed as a reliability cut-off.

Patients with a SLD ≤2.5cm showed considerably higher accuracy in differentiating between histopathology Metavir F scores, particularly at lower F score cut-offs (Table 4.27). In regards to identifying clinically significant discordance between ARFI and liver histology (≥2 F score difference), applying a SLD cut-off 2.5cm provided a sensitivity of 48.0% (95%CI: 32.0 – 65.2%) and a specificity of 86.8% (95%CI: 77.3 – 92.9%) in detecting these inaccurate studies.

Table 4.26. Sensitivity, specificity, PPV and NPV of ARFI in diagnosing significant fibrosis (≥F2), severe fibrosis (≥F3) and cirrhosis (F4) amongst all patients (first row), patients passing IQR/Median criteria (second row) and patients with a SLD ≤2.5cm (third row). Applying IQR/Median and SLD criteria improved overall accuracy (particularly specificity), with patients 'passing' SLD criteria generally showing higher accuracy than those 'passing' IQR/Median criteria.

F score cut-off	Accuracy (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)
F0/F1 vs. ≥F2					
All patients	60.7% (52.8 - 68.2%)	80.4% (67.4 – 89.2%)	42.9% (20.8 – 55.9%)	56.2% (44.8 – 70.0%)	70.6% (53.7 – 83.3%)
Pass IQR/Median Criteria	59.3% (48.7 - 69.1%)	73.0% (56.9 – 84.8%)	49.0% (35.6 – 26.5%)	52.0% (38.7 – 64.9%)	70.6% (53.7 – 83.3%)
SLD ≤2.5cm	68.3% (57.6 – 77.4%)	82.9% (68.4 – 91.8%)	53.7% (38.7 – 68.0%)	64.2% (50.7 – 75.7%)	75.9% (57.6 – 88.1%)
≤F2 vs. ≥F3					
All patients	67.3% (57.5 – 74.3%)	88.9% (71.1 – 97.0%)	60.0% (49.0 – 70.1%)	42.9% (30.8 – 55.9%)	94.1% (82.5 – 98.6%)
Pass IQR/Median Criteria	74.4% (64.2 – 82.5%)	84.2% (61.6 – 95.3%	71.6% (59.9 – 81.1%)	45.7% (30.4 – 61.8%)	94.1% (83.5 – 98.6%)
SLD ≤2.5cm	76.8% (66.6 – 84.7%)	87.5% (68.2 – 96.5%)	72.4% (59.7 – 82.3%)	56.8% (40.9 – 71.3%)	93.3% (81.5 – 98.4%)
≤F3 vs. F4					
All patients	68.2% (58.9 – 76.3%)	80.0% (54.1 – 93.7%)	66.3% (56.1 – 75.2%)	27.9% (16.6 – 42.8%))	95.3% (86.6 – 98.9%)
Pass IQR/Median Criteria	75.6% (65.5 – 83.5%)	77.8% (44.3 – 94.7%)	75.3% (64.6 – 83.7%)	26.9% (13.5 – 46.3%)	96.7% (88.0 – 99.8%)
SLD ≤2.5cm	76.8% (66.6 – 84.7%)	78.6% (51.7 – 93.2%)	76.5% (65.1 – 85.1%)	40.7% (24.5 – 59.3%)	94.5% (84.6 – 98.7%)

Table 4.27: Accuracy of ARFI at differentiating between histopathology Metavir F scores at the F1/2, F2/3 and F3/F4 cut-offs.

SLD	AUROC at F1/2 (95%CI)	AUROC at F2/3 (95%CI)	AUROC at F3/4 (95%CI)
≤2.5cm	0.79 (0.69 – 0.89)	0.84 (0.74 – 0.94)	0.76 (0.62 – 0.90)
>2.5 cm	0.41 (0.17 – 0.64)	0.64 (0.41 – 0.86)	0.79 (0.63 – 0.96)

When combined with IQR/Median criteria, applying an SLD cut-off of ≤2.5cm provided better stratification of inter-operator concordance and accuracy than either marker alone (Table 4.28, Figure 4.8). In particular, patients who passed both IQR/Median and SLD criteria showed significantly better correlation with histopathology (rho = 0.487 vs. rho = 0.008) and better concordance with Metavir F score (84.7% vs. 45.9%, Fisher' exact, p=0.001) than those failing either or both criteria. Patients who passed both SLD and IQR/Median criteria (suggestive of more reliable / accurate ARFI results) encompassed the majority (i.e. 70.4%) of the greater patient cohort.

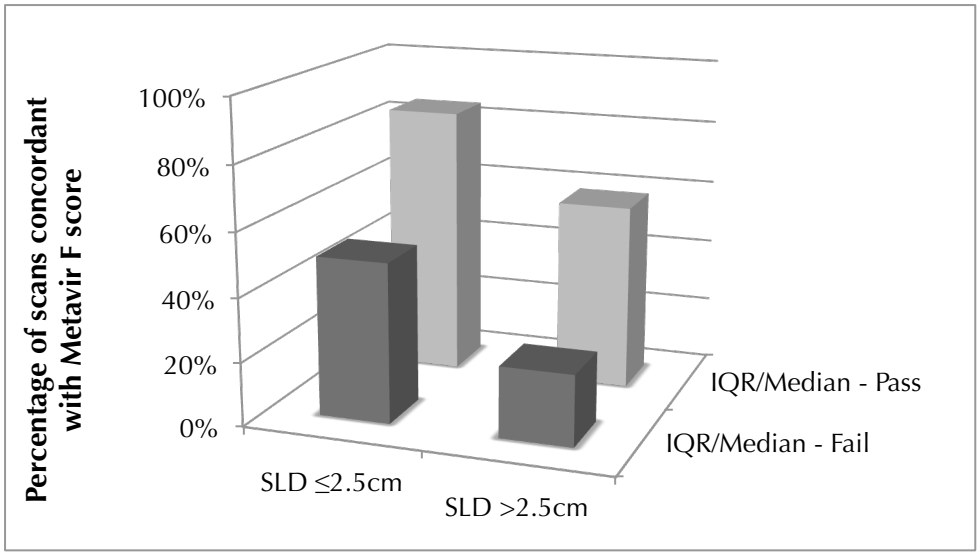
In multi-regression analyses, the amount ARFI velocities deviated from the histology reference range shown a much stronger correlation with SLD (R^2 = 0.508) than with IQR/Median ratio (R^2 = 0.071). This further suggests SLD is more closely associated with ARFI accuracy than is IQR/Median.

Table 4.28: Utility of combining IQR/Median and SLD reliability criteria. The two criteria appeared to show better stratification of ARFI accuracy than either criteria alone.

Predictive Marker	% ARFI sets discordant with biopsy	Correlation between ARFI velocity and Biopsy F score (rho)
IQR/Median criteria alone		
Pass	14.82% (95%CI: 7.4–26.9%)	R = 0.291
Fail	35.71% (95%CI: 25.5-47.4%)	R = -0.146

SLD criteria alone		
Pass (SLD≤2.5cm)	19.59% (95%CI: 12.8-28.7%)	R = 0.516
Fail (SLD>2.5cm)	51.85% (95%CI: 34.0-69.3%)	R = -0.188
IQR/SR & SLD combined		
Pass IQR/Median & Pass SLD	15.28% (95%CI: 8.58-25.50 %)	R = 0.487
Fail IQR/Median & Pass SLD	50.00% (95%CI: 28.00-72.00%)	R = 0.230
Pass IQR/Median & Fail SLD	41.67% (95%CI: 19.26-68.11%)	R = -0.153
Fail IQR/Median & Fail SLD	77.78% (95%CI: 44.3-94.66%)	R = -0.598

Figure 4.8: Percentage of ARFI scans being concordant with liver histopathology, amongst those passing/failing SLD and IQR/Median criteria.



As previously mentioned, there was a low spread in SLD values observed between the 10 measurements taken within each individual patient. Measuring the SLD from the first ARFI measurement provided an estimation within 0.07cm from the set's overall SLD value, and correctly stratified patients into SLD ≤2.5 vs >2.5cm categories in nearly 95% of cases. The preciseness of the SLD estimate achieved after measuring the SLD from an increasing number of ARFI measurements is outlined in Table 4.29 below.

Table 4.29: Precision of the SLD estimation (deviation from the median SLD of 10 measurements) achieved by measuring the SLD from an increasing number of ARFI readings. Measuring a single ARFI reading provided a close SLD estimation to the set's overall median.

No of readings in the set measured	Deviation of measured SLD (from the set's overall SLD)	% of patients correctly stratified into SLD ≤2.5 vs >2.5cm groups
1	0.69mm	94.4% (95%CI: 92.6–95.8%)
2	0.51mm	95.4% (95%CI: 93.7-96.7%)
3	0.43mm	96.4% (95%CI: 94.9-97.5%)
4	0.35mm	96.9% (95%CI: 95.4-97.9%)
5	0.27mm	97.8% (95%CI: 96.5-98.6%)
6	0.22mm	98.6% (95%CI: 97.5-99.3%)
7	0.16mm	98.8% (95%CI: 97.7-99.4%)
8	0.12mm	99.1% (95%CI: 98.2-99.7%)
9	0.07mm	99.1% (95%CI: 98.2-99.7%)

4.6.3 Inter-operator agreement

We also assessed whether routinely scanning patients with multiple independent operators (i.e. looking at inter-operator concordance) had value in triaging the reliability of ARFI measurements. When patients were scanned with two operators, those with concordant ARFI readings (i.e. within adjacent F scores) showed considerably better correlation with histopathology (rho = 0.392) than those with discordant operators (rho = 0.010). The assessment of ARFI reliability was further improved by assessing patients with three independent operators, with ARFI having an even stronger correlation with histopathology (rho = 0.571) in cases of three-way inter-operator concordance. These findings are further outlined in Table 4.30.

Table 4.30: Level of concordance / correlation between ARFI and biopsy, according to whether the patient's operators had concordant vs. discordant ARFI F scores.

Number of operators concordant	% Concordant with Histopathology	Correlation between ARFI velocity and histology
Two Operators		
Operators concordant	71.4% (61.0 – 80.0%)	Rho = 0.392
Operators discordant	65.0% (43.2 – 82.0%)	Rho = 0.010
Three Operators		
Operators concordant	82.2% (68.4 – 91.0%)	Rho = 0.571
Operators discordant	66.7% (45.2 – 83.0%)	Rho = 0.180

Similar findings were observed when the percentage deviation between operators was analysed (Table 4.31). Those patients in whom there was close agreement between the first two operator ARFI velocities (<10% deviation) had ARFI values which closely approximated the histopathology reference (mean deviation = 0.19 m/s). Conversely, those with significant inter-operator deviation (i.e. >20%), deviated a mean of 0.60 m/s from the histopathology F score. This difference was statistically significant (unpaired t test, $p<0.003$).

Amongst the small number of patients in the biopsy sub-cohort with discordant operator F scores (n=11), the lower of the two operator ARFI values appeared to be the more accurate. The lower operator's ARFI velocity deviated a mean of 0.167 m/s from the histology reference range, compared to 0.427 m/s for the operator with the higher ARFI velocity (unpaired t test = 0.09). The lower operator's ARFI velocity also appeared more accurate than the mean of the two operator's ARFI velocities (Table 4.32). On average, the lower operator's ARFI velocity fell slightly below the histology F score range (mean = -0.167 m/s), whilst the higher operator's ARFI velocity was fell significant above the histopathology reference (mean = +0.427 m/s).

Table 4.31: The level of inter-operator agreement (i.e. percentage deviation between operator velocities) correlated with the degree of deviation between ARFI and histopathology. Namely, patients in whom both operators had concordant ARFI scores (i.e. <10% of each other) showed better accuracy than those with poorer operator agreement (>20% deviation).

Percentage deviation between operators	Number of patients (%)	Mean deviation (m/s) from the histopathology reference F score range
<10%	34 (33%)	0.188 m/s
10 to 20%	16 (15%)	0.371 m/s
20 to 40%	28 (27%)	0.629 m/s
>40%	26 (25%)	0.557 m/s

Table 4.32: Mean net and gross deviation of operator ARFI velocities from the histology reference range, amongst the small number of patients in whom operator ARFI F scores were discordant (n=11). In these cases, the lower ARFI operator velocity appeared to be more accurate, and also appeared to more closely approximate histopathology than did taking the mean of the two operator's values.

Operator Value	Mean net deviation from the biopsy F score range (m/s)	Mean gross deviation from the biopsy F score range (m/s)
Lower Operator ARFI velocity	-0.167 m/s	0.167 m/s
Higher Operator ARFI velocity	0.427 m/s	0.427 m/s
Mean of the two operators' ARFI velocities	0.146 m/s	0.284 m/s

We also assessed which factors predicted the presence of inter-operator discordance, with a view to identifying which patients will be most likely to benefit from scanning with

additional operators in clinical practice. In multi-regression analyses (Table 4.33), the factors which showed the strongest independent association with inter-operator deviation included SLD (R2 = 0.19), IQR/Median ratio (R2 = 0.19) and to a lesser extent operator ARFI F score (R2 = 0.10). The impact of other factors include age, gender, steatosis and BMI was minimal / non-independent. It would therefore appear that scanning patients with multiple independent operators would have greatest yield amongst patients with high SLD values, high IQR/Median ratios or high ARFI F scores; as inter-operator discordance is most frequent in these patients.

Table 4.33: Multi-regression analyses looking at the association between patient factors and the percentage between operators within each patient. SLD, IQR/Median ratio and ARFI F score all showed the strongest independent associations with inter-operator discordance.

Patient Variable	Association between patient factor and % deviation between operators
SLD (cm)	$R^2 = 0.19$
IQR/Median Ratio	$R^2 = 0.19$
ARFI F score	$R^2 = 0.10$
Hepatosteatosis	$R^2 = 0.05$
Gender	$R^2 = 0.05$
Age	$R^2 = 0.04$
BMI	$R^2 = -0.04$

Applying binary cut-offs of IQR/Median >0.3, SLD >2.5cm or ARFI F score of F3/F4 provided good stratification of the likelihood of inter-operator discordance (Table 4.34). Patients with either an IQR /Median >0.3 or a SLD>2.5cm had the highest rates of inter-operator discordance (23.1%, 95%CI: 18.8 – 28.0%). Patient with an IQR/Median ≤0.3, SLD ≤2.5cm and F3/F4 fibrosis on ARFI had intermediate rates of inter-operator discordance (12.6%, 95%CI: 8.1 – 18.9%). Whilst those with an IQR/Median ≤0.3, SLD ≤2.5cm and F0-F2 fibrosis on ARFI had very low rates of inter-operator discordance (6.2%, 95%CI: 4.3 – 9.8%).

Table 4.34: Inter-operator discordance rates according to IQR/Median ratio, SLD value and ARFI fibrosis score.

	Inter-operator discordance rate (% and 95%CI)
IQR/Median ≤0.3	9.65% (7.67 – 12.06%)
IQR/Median >0.3	23.89% (18.78 – 29.88%)
SLD ≤2.5cm	10.40% (8.37 – 12.85%)
SLD > 2.5cm	22.38% (17.25 – 28.51%)
ARFI - F0/F1/F2	8.87% (6.79 – 11.49%)
ARFI - F3/F4	19.94% (16.69 – 23.66%)

4.7 Cirrhosis assessment

4.7.1 Child-Pugh and MELD scores

A moderate positive correlation was seen between ARFI velocity and MELD score (rho = 0.342, p=0.0001); indicating overall higher ARFI readings with increasing MELD score severity. A wide range in ARFI velocities was nonetheless seen across all MELD score levels, which is illustrated in Figure 4.9. As a consequence, ARFI had only a modest AUROC of 0.67 (95%CI: 0.57 – 0.76) in discriminating early (i.e. MELD score ≤10) vs. more advanced cirrhosis (MELD score >10).

Similar findings were observed for Child-Pugh, with ARFI velocity showing a moderate positive correlation with the Child-Pugh Score (rho = 0.363, p<0.0001, Figure 4.10). The mean ARFI velocity amongst patients with Child-Pugh A vs. B/C cirrhosis was 2.36 vs. 2.82m/s respectively (unpaired t test, p=0.0005). ARFI had an AUROC of 0.713 (95%CI:

0.62-0.81) in discriminating between patients with Child-Pugh A vs. B/C cirrhosis.

Figure 4.9: Relationship between ARFI velocity and MELD score, showing a moderate positive correlation between the two variables.

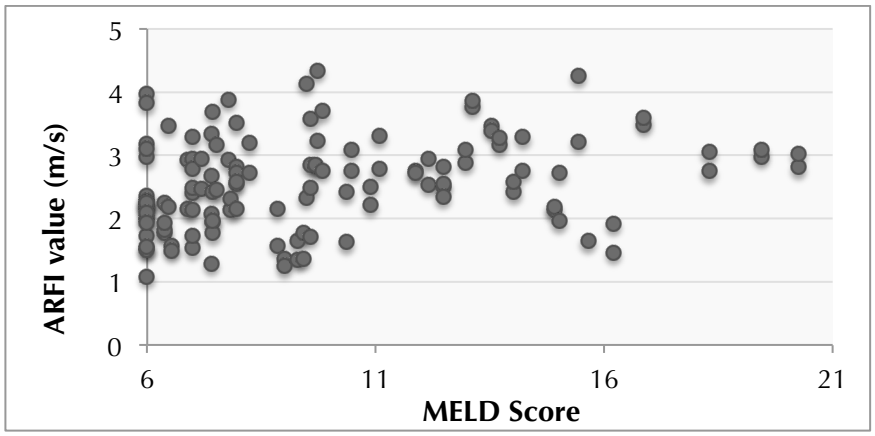
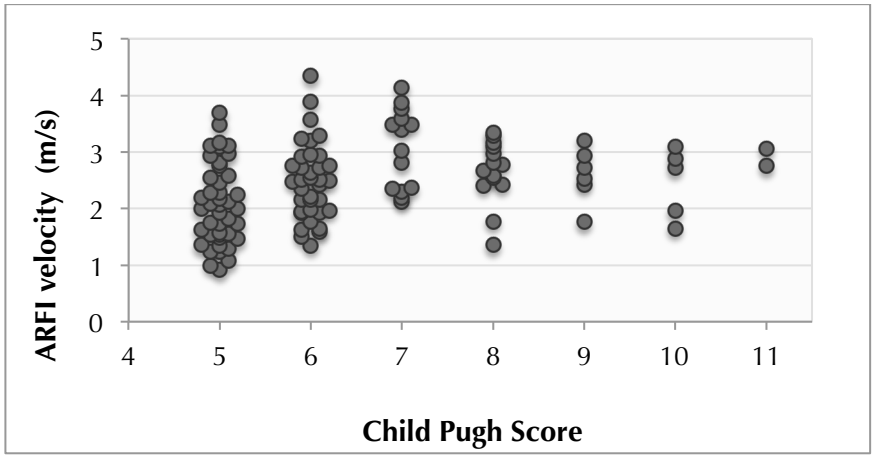


Figure 4.10: Relationship between ARFI velocity and Child-Pugh Score. Whilst there was a moderate positive correlation (rho = 0.363, p<0.0001) between the two variables, significant overlap was seen between the ARFI velocities of the respective Child-Pugh Scores.



4.7.2 Blood parameters

ARFI also showed a statistically significant, albeit modest relationship with blood test parameters linked with hepatic reserve and portal hypertension. Increasing ARFI LSMs were associated with lower levels of albumin (rho = -0.2098, p = 0.001) and platelets (rho

= -0.124, p=0.05), increasing bilirubin levels (rho = 0.194, p = 0.03) and increased INR (rho = 0.230, p = 0.004).

4.7.3 Cirrhotic complications

ARFI showed only a weak association with the presence or absence of cirrhotic complications and decompensation (summarised in Table 4.35). Slightly higher ARFI velocities were however observed amongst patients with portal hypertension (mean ARFI velocity = 2.61 vs. 2.45m/s, unpaired t test, p=0.04), hepatic encephalopathy of any severity (mean = 2.76 vs. 2.49m/s, unpaired t-test, p=0.03) or ascites (mean = 2.70 vs. 2.48m/s, unpaired t test, p=0.04). The predictive power (i.e. AUROC) of ARFI in discriminating between patients with or without portal hypertension, any encephalopathy and any ascites was 0.542, 0.602, 0.580 respectively.

ARFI showed the strongest predictive power for oeosphageal varices, demonstrating an AUROC of 0.687 (95%CI: 0.593 – 0.781) for detecting any varices, and 0.725 (95%CI: 0.465 – 0.785) for detecting medium/large varices. The mean ARFI velocity amongst patients with or without varices was 2.67 and 2.30m/s, respectively (unpaired t test, p=0.004). Significant overlap was however again seen between the ARFI velocities of patients with no, small or medium/large varices, as demonstrated in Figure 4.11.

Figure 4.11: ARFI velocities amongst patients with no, small or medium/large varices at gastroscopy.

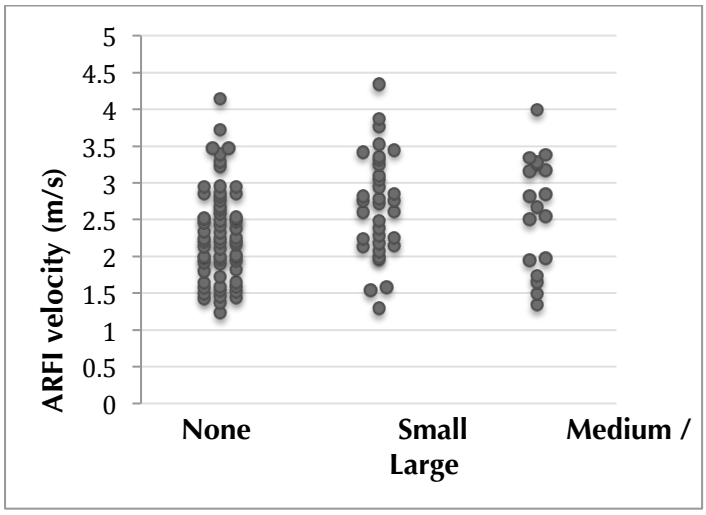


Table 4.35: Relationship between ARFI velocity and the presence complications resulting from portal hypertension or synthetic failure. ARFI showed only modest predictive power (i.e. AUROC) for detecting the presence of cirrhotic complications.

	Median ARFI velocity (m/s) (Quartile 1, Quartile 3)	**AUROC (95% CI)**
Portal hypertension		
No	2.46 (1.96, 3.17)	0.542
Yes	2.71 (2.17, 3.09)	(0.484 – 0.601)
Oesophageal varices		
		Any Varix:
No	2.24 (1.86, 2.66)	0.687 (0.593 – 0.781)
Small	2.74 (2.20, 3.21)	
Medium / Large	2.75 (1.96, 3.23)	Med/Large Varices:
		0.725 (0.465 – 0.785)
Encephalopathy		
		Any Encephalopathy:
No	2.49 (1.98, 3.03)	0.602 (0.511 – 0.694)
Grade I / II	2.76 (2.21, 3.28)	
Grade III / IV	2.76 (2.42, 3.12)	Grade III / IV:
		0.598 (0.481 – 0.715)
Ascites		
		Any ascites:
No	2.48 (1.96, 3.06)	0.580 (0.508 – 0.651)
Mild	2.75 (2.44, 3.16)	
Moderate / Severe	2.54 (2.25, 2.90)	Moderate / Severe:
		0.527 (0.405 – 0.649)

Even amongst the most reliable measurements (i.e. those with a SLD <2.5cm and passing IQR/Median criteria), ARFI still showed only modest accuracy in predicting the presence of varices; having an AUROC of 0.700 (95%CI: 0.569-0.830) in detecting any varices and an AUROC of 0.547 (95%CI: 0.347 – 0.747) for detecting medium/large varices.

Chapter 5. Discussion – ARFI (Siemens)

5.1 Fibrosis quantification

5.1.1 ARFI Accuracy

ARFI has been almost exclusively validated in European and Asian cohorts to date, and our study provides one of the first indications of local ARFI performance in a clinical Australian setting. In summary, we found ARFI to have moderate local accuracy in the quantification of liver fibrosis, demonstrating an AUROC of 0.67, 0.76 and 0.70 at diagnosing significant fibrosis (≥F2), severe fibrosis (≥F3) and cirrhosis (F4), respectively. ARFI showed high sensitivity (80.0 – 88.9%) and NPV (70.6 – 95.3%), but relatively poor specificity (42.9 – 66.3%) and PPV (27.9 – 56.2%) at differentiating the three F score cut-offs.

ARFI's primary value in clinical practice therefore appears to be in the exclusion of liver fibrosis. The tool's high sensitivity and NPV at the three F score cut-offs means a negative ARFI result is reassuring that a patient is likely to have either no or low levels of liver fibrosis. ARFI is therefore likely to have greatest utility as an initial screening tool; helping to identify patients with a low likelihood of liver fibrosis in whom further investigation or treatment may not be required.

In contrast, ARFI showed a propensity towards false positive results in our cohort, which is reflected in the tool's modest specificity and PPV at the three F score cut-offs. Whilst these findings require further confirmation (in view of the study limitations outlined below), they nonetheless caution against placing heavy reliance on elevated ARFI results in clinical practice. Current elastography guidelines recommend that a diagnosis of advanced fibrosis can be made on the basis of an elevated ARFI LSM alone.[35,140] Our results, however, suggest this practice may not be locally advisable. Elevated ARFI LSM values should instead be interpreted with care, and with knowledge and attention to any potential confounding variables which may provide an alternative explanation for the result. Our findings would also suggest that any diagnosis of advanced fibrosis should ideally be supported by supplementary evidence, be this from clinical history, serologic fibrosis markers or even liver biopsy in reserved cases. Our cautionary findings nonetheless require confirmation in a larger local cohort, given their conflict with existing guidelines.

The observed ARFI accuracy was also lower than is commonly described in the literature. In the widely cited meta-analysis by Friedrich-Rust *et al.,* ARFI showed an accuracy (i.e. AUROC) of 0.87, 0.91 and 0.93 in detecting ≥F2, ≥F3 and F4 disease, respectively, in the setting of HCV.[145] Similar results have also been reported in a larger meta-analysis encompassing a range of CLD aetiologies (36 studies, 3951 patients), which showed an overall AUROC of 0.84, 0.89 and 0.91 at diagnosing ≥F2, ≥F3 and F4 disease.[205] Another meta-analysis by Bota *et al.* reported a summary sensitivity / specificity of 0.74 / 0.83 for the diagnosis of significant fibrosis (≥F2), and 0.87 / 0.87 for the diagnosis of cirrhosis (F4).[15] Our results therefore suggest local ARFI accuracy may be lower than widely reported, driven by the apparent reduction in local specificity.

Whilst our study limitations may result in the underestimation of local ARFI accuracy, our findings are likely to be at least partially explained by a true local reduction in ARFI performance. We found ARFI reliability to be reduced throughout the greater patient cohort, rather than being limited to the biopsy sub-cohort. The greater cohort showed higher median IQR/Median ratios and higher failure rates than is commonly reported; with 15.5% of the cohort 'failing' IQR/Median criteria compared to only 6.1% of scans in other studies.[11] We also found significant rates of inter-operator discordance, at levels greater than would be expected from the very high inter-operator reproducibility previously reported.[158–161] It would therefore be expected that local accuracy would be similarly reduced.

There are a number of possible explanations for the likely true reduction in local ARFI performance. Our cohort firstly had a high obesity rate, with 32.2% of patients being overweight (BMI = 25 – 30kg/m^2) and 25.8% obese (BMI >30kg/m^2). This rate is considerably higher than the majority of published studies, which reflects the high obesity rates in Australia relative to Europe or Asia. This population difference is significant, given the marked impact of obesity on ARFI performance in our study (outlined below). An additional factor may include operator training, with operators in our study having a highly variable levels of point SWE experience. Alternatively, the results may simply reflect differences in ARFI performance when performed in a 'real world' clinical setting, external to rigors of research protocols.

5.1.2 ARFI F score cut-offs

The finding of high sensitivity / NPV but low specificity / PPV may also be partially attributable to the ARFI LSM cut-offs selected for clinical application. The adopted F score cut-offs were drawn from a meta-analysis by Friedrich-Rust *et al.,* which pooled the results of 518 patients from 8 different studies evaluating ARFI in the setting of HCV.[145] It has been demonstrated that slightly different ARFI velocities are observed amongst patients with different aetiologies of CLD, and the F score cut-offs described by Friedrich-Rust *et al*. are lower than those reported in many studies. Applying these relatively lower cut-offs may therefore have contributed to the high sensitivity but low specificity observed in our own cohort.

After applying cut-offs which had been statistically optimized to our own cohort (i.e. higher than those of Friedrich-Rust *et al.*), we did see some improvement in overall ARFI accuracy, specificity and PPV (Table 4.9). Specificity and PPV nonetheless remained suboptimal, ranging between 64.3-72.5% and 28.2-62.3% at the three F score cut-offs, respectively, and the choice in F score cut-offs is therefore unlikely to be the predominant explanation for the technology's significant false positive rate. The improvement in specificity / PPV with the optimized F score cut-offs came predictably, however, at the expense of reduced sensitivity and NPV (Table 4.9).

Both sets of cut-offs therefore have their pros and cons, and arguments could be made for their differential application depending on the clinical goal. More conservative cut-offs (i.e. weighting sensitivity over specificity) may however be preferable for most clinical settings, to help ensure the highest number of patients with liver fibrosis are diagnosed and therefore receive appropriate treatment. The continued application of the cut-offs drawn from Friedrich-Rust *et al.* may therefore be appropriate, acknowledging that this is at the expense of increased false positive results.

5.1.3 Limitations of accuracy analyses

Our accuracy analyses do however have a number of limitations and potential biases, which may have contributed to the low ARFI accuracy observed. The first includes possible referral bias resulting from the small proportion of patients completing a contemporaneous liver biopsy. This reflects trends in local clinical practice, with non-

invasive assessment tools now replacing liver biopsy for a large number of clinical indications. Less than 6% (n=55) of the patient cohort had contemporaneous liver histopathology available for comparison, which severely reduced the power of our accuracy analyses. It also introduces the risk of referral bias, with the group of patients completing a liver biopsy potentially being non-representative of the greater ARFI cohort. Of particular note is the high rate of autoimmune hepatitis and necroinflammatory change observed in the liver biopsy sub-cohort, which has the potential to falsely elevate LSMs and thereby contribute to the high rate of false positive results observed. This is likely to have contributed to the lower observed accuracy findings, with the AUROC significantly improving amongst patients without necroinflammatory change on biopsy (Table 4.15). The other consideration is that patients may have been referred for liver biopsy in response to an unexpected ARFI result. On reviewing only patients scanned on the day of or following liver biopsy (n=35), the correlation between ARFI velocity and Metavir F score was not, however, significantly changed.

Liver biopsy also represents an imperfect gold standard for fibrosis assessment. Irrespective of the accuracy of a non-invasive tool, using an imperfect reference will invariably result in a lower perceived accuracy. This may be further exacerbated by the use of a portal tract cut-off of ≥6 in our study. Whilst this cut-off is widely employed in other studies,[15] some guidelines do recommend the use of a more stringent minimum cut-off of 11 portal tracts to minimize sampling error.[71] We furthermore applied a cut-off of six months time difference between the ARFI and liver biopsy for our analyses. This threshold has been used by a number of published studies, however it too introduces further potential for error. Whilst adopting stricter biopsy inclusion requirements would be ideal, the already small size of the biopsy cohort precluded such measures being taken.

As a consequence of these limitations, our accuracy analyses are not designed or intended to provide a precise representation of local accuracy, and they likely underestimate ARFI's true accuracy in a clinical Australian setting. Some inferences can nonetheless be drawn from our results; namely that ARFI is an imperfect tool which can be associated with false positive results in a significant proportion of patients. Most importantly, however, the accuracy results provide a platform from which a wide range of sub-analyses can be performed. Whilst the above limitations need to be kept in mind when interpreting the absolute accuracies of the below sub-analyses, the biases apply to the whole biopsy sub-cohort and are therefore unlikely to explain any relative differences in accuracies described below.

5.2 Body habitus

5.2.1 Study findings

We found body habitus to have a strong relationship with all facets of ARFI performance. ARFI reliability appeared to be particularly reduced in the setting of obesity (BMI >30kg/m^2), with obese patients showing considerably higher median IQR/Median ratios, greater inter-operator and poorer correlation with biopsy when extrapolating from SLD data. These findings were not only of strong statistically significance, but also likely to be of high clinical significance given the effect sizes observed.

The other notable finding was the propensity towards falsely elevated LSMs amongst obese patients. Measurements with a SLD >2.5cm (equivalent to a BMI of >30kg/m^2) showed greater deviation above the biopsy reference F score range than measurements with a SLD ≤2.5cm (mean net deviation = + 0.951 vs. 0.214 m/s, p=0.001). This was further confirmed in multi-regression analyses, which showed a strong positive relationship between body habitus and ARFI velocity that was independent of liver fibrosis severity. These findings strongly caution against relying on elevated ARFI LSM values in the setting of obesity, and also provide a possible explanation for the low specificity / PPV observed in the overall accuracy sub-cohort.

5.2.2 Comparison with existing literature

Our results are supported by a number of studies which have implicated obesity in the reduction in ARFI reliability. ARFI was initially favoured over TE in the setting of obesity, due its ability to obtain measurements in the vast majority of patients, irrespective of BMI. There is, however, increasing awareness that obesity can impact on all facets of ARFI performance, being linked with increased failure rates of IQR/Median reliability criteria,[11] reduced intra & inter-observer reliability[158] as well as lower accuracy.[23,24,169]

Despite the increasing body of evidence, there has been inconsistency between studies. The magnitude of obesity's impact on ARFI reliability has been highly variable, and some groups have reported body habitus to have no significant association with ARFI performance.[28,47] As an example, Attia *et al.* assessed ARFI accuracy amongst a cohort of exclusively overweight (BMI = 25 – 30kg/m^2, n=61) or obese patients (BMI = >30kg/m^2,

n=26). They found ARFI to have an AUROC of 0.94 and 0.97 in detecting cirrhosis in the two BMI brackets, respectively, and concluded that ARFI performance was therefore not significantly degraded by increased body habitus.[28]

Our findings are however much more cautionary. The magnitude of obesity's impact on ARFI performance was firstly greater than has been observed in the majority of studies mentioned above. Our results are also drawn from one of the largest patient cohorts evaluated with ARFI, which also has a high rate of obesity. Our study therefore provides one of the strongest warnings regarding the importance of body habitus on ARFI performance to date.

5.2.3 Clinical implications

Our obesity findings have important implications for clinical practice. Clinicians need to be cautious in interpreting and relying on ARFI results in obese patients, particularly in the case of elevated LSMs (i.e. F3 / F4 readings). Whilst a low ARFI velocity remains reassuring that a patient likely has low levels of fibrosis, an elevated LSM has limited value and should be regarded as a prompt for further investigation rather than an indication of advanced fibrosis.

This limitation is significant, as it reduces ARFI's utility in the clinical assessment of NAFLD, which frequently co-exists with obesity. NAFLD itself covers a wide spectrum of diseases ranging from simple hepatosteatosis, steatohepatitis through to cirrhosis.[206] It is therefore essential that clinicians can accurately screen this large population of patients, to identify those at risk who require closer monitoring and possibly intervention. This is now regarded as one of the clinical priorities in hepatology, with the need becoming increasingly heightened by the recent 'NAFLD epidemic'. Other common elastography tools, particularly TE, have limited application in population screening due to their propensity towards false positive results in diabetic and obese patients.[207,208] Our results suggest ARFI is likely to have a similar limitation in this clinical setting.

5.2.4 Skin-to-Liver Capsule Distance (SLD)

The impact of body habitus on ARFI performance is widely attributed to the attenuation of the ARFI push pulse from increasing depths of subcutaneous adipose tissue (discussed in detail below). We therefore aimed to assess whether the subcutaneous depth (i.e. SLD) is more intimately related with ARFI performance than is BMI, and our results suggest this is the case. We found SLD to have a slightly stronger correlation with IQR/Median and inter-operator deviation than did BMI, and SLD also appeared to have a stronger association with ARFI performance in multi-regression analyses. Whilst further analysis of the relative impact of SLD vs. BMI on ARFI accuracy is required, our findings suggest subcutaneous tissue depth and central adiposity are more closely related to ARFI performance than is overall body habitus.

The literature surrounding SLD and ARFI is limited, with only two studies to our knowledge evaluating the relationship between SLD and ARFI performance. The first by Cassinotto *et al.* observed a higher proportion of measurements to fail IQR/Median criteria with increasing 'parietal wall' thickness.[24] A second smaller study by Karlas *et al.* evaluated the role of SLD amongst 41 patients with morbid obesity undergoing low-energy diets prior to bariatric surgery.[169] They observed ARFI accuracy to be reduced amongst patients with an SLD >3.58cm and >3.33cm, before and after diet intervention respectively. The impact of SLD on ARFI performance at lower levels of obesity is however less well established, and the relative impact of SLD vs. BMI on ARFI performance has not been comprehensively evaluated to our knowledge.

There is also early evidence that SLD may hold relevance for other elastography tools. A small number of studies have found increased SLD values to be associated with higher rates of unsuccessful measurements in the context of SSI.[24,183,209] Skin-to-liver Capsule Distance is also widely recognized in TE, being commonly referred to as 'SCD' in the literature. Fibroscan® reliability has shown to reduce once the SCD exceeds 2.5cm, and use of the XL probe is therefore recommended above this cut-off.[210] Whilst it is possible that ultrasound beam attenuation may contribute to the reduction in Fibroscan® performance with increasing SCD, the primary factor is likely to be the M probe's fixed focal length of 2.5 to 5cm; which may result in measurements being inadvertently acquired over the liver capsule / subcutaneous tissue in obese patients. The differences in the likely underlying mechanism limits the extrapolation of the more extensive Fibroscan® SCD literature to either point SWE or 2D-SWE devices.

5.2.5 Mechanism

Our results also provide some insight into the mechanism underlying body habitus' impact on ARFI reliability. The relative impact of SLD versus BMI on ARFI performance firstly supports the current working hypothesis implicating subcutaneous adipose tissue in the attenuation and degradation of the ultrasound beam.[148] The effect of adipose tissue on ultrasound is well recognized, with the best example being beam attenuation from hepatosteatosis in B-mode imaging. The researchers from Duke University who originally developed ARFI technology speculated that similar reliability issues may also apply for ARFI.[170] They hypothesized a potential for fat to interfere with the tracking of tissue displacement, but also to potentially reduce the amplitude of the high-energy acoustic impulse. The relative impact of SLD versus BMI on ARFI performance would therefore support this hypothesis.

Some inferences regarding the impact of beam attenuation on ARFI LSMs can also be drawn from phantom experiments by Chang *et al*. and Kaminuma *et al*.[21,166] Both studies found ARFI performance (i.e. measurement reproducibility) to worsen once the total measurement depth exceeded 8cm. These 'deep' measurements were also associated with lower ARFI velocities, which was hypothesized to reflect increased attenuation of the ultrasound push pulse, reduced local energy transfer and thereby lower generated shear wave velocities. Our results relating to SLD are both concordant and discordant with these phantom model experiments. Whilst we observed reduced measurement reproducibility with increasing SLD, central adiposity was associated with falsely elevated rather than reduced shear wave velocities. This is somewhat unexpected, as one might expect ARFI LSMs to be similarly reduced as a result of ultrasound attenuation. This discrepancy raises the possibility of additional mechanisms underlying the relationship between SLD and ARFI reliability, beyond simply the attenuation of the push pulse. Possible explanations could include degradation of the tracking ultrasound beams, which are of higher frequency therefore inherently more prone to attenuation. Alternatively, fat is associated with increased ultrasound scatter, which theoretically may cause dispersion of the push pulse and thereby possibly interfere with shear wave generation and tracking.

Another explanation is that central adiposity may be interfering with the technical ability of operators to reliably acquire ARFI measurements. This hypothesis is raised by the study from Gradinaru-Toscau *et al.,* who tested 371 CLD patients with SSI by both novice and expert operators.[188] They found both operators to show similar rates of reliable

measurements amongst patients of normal BMI (92.3% vs. 97.5%, p=0.24), however found the operators to have increasingly disparate performance in the setting of obesity; reliable measurements being observed in 45.9% vs. 73.4% of patients scanned by the novice vs. expert operator, respectively (p=0.03). The study concluded the degradation in SWE performance with obesity is likely to be operator dependent, which suggests other operator technical factors may be involved beyond purely beam attenuation. Whilst such analyses have not been replicated in ARFI, the findings are likely to be applicable to point SWE given the shared technology underpinning the techniques. One possible operator technical factor could include difficulties in successfully angling the ultrasound beam through the intercostal space due to increasing depths of intervening tissue, potentially causing the push pulse and tracking beam to be partially reflected by the ribs if not angled correctly. Such a mechanism could explain some of the discrepancies observed in our data. This includes the observation of paradoxically elevated ARFI LSM with increasing SLD, the highly variable measurement velocities obtained within each patient (one may expect attenuation to have a more uniform impact on measurements), and finally the seemingly disparate findings between studies; which would again point towards the presence of operator dependent factors.

5.2.6 Limitations

Our body habitus findings do have limitations, which primarily relate to the small size and potential biases of our biopsy cohort. Whilst the marked impact of body habitus on ARFI reliability (particularly IQR/Median and inter-operator concordance) is unequivocal, our results may not provide a true indication of body habitus' impact on ARFI accuracy. This would require further evaluation in a larger prospective study with a histopathologic correlate.

Secondly, our study did not control for ultrasound probe pressure, which could theoretically impact on measurement SLDs as a result of variable compression. Nonetheless, the effect of probe pressure on SLD is felt likely to be negligible. We observed minimal variation (median deviation = 5.9%) between the SLD of measurements obtained by different operators in each patient. Our results therefore suggest variations in operator technique (including probe pressure) has limited influence on the SLD value recorded in individual patients.

Our results do not shed light on whether increases in SLD caused by ascites have a similarly negative impact on ARFI reliability. If SLD's impact is attributable to beam attenuation from increasing adipose tissue or technical considerations from increased distance from the intercostal spaces, as hypothesized above, then any increase in SLD due to ascites may not necessarily reduce ARFI accuracy. This hypothesis would however require further assessment in future studies.

Finally, it is unclear whether operators should modify their acquisition technique (i.e. the intercostal approach) to help reduce SLD and thereby minimize ultrasound beam attenuation. Given the potential biases involved, this question would need to be formally assessed in a randomized clinical trial.

5.3 Additional patient factors affecting ARFI reliability

5.3.1 Necroinflammatory change

Hepatic inflammation, as indicated by an ALT level >100 IU/L or a Metavir A score ≥2, was found to be independently associated with reduced ARFI reliability in our cohort. Necroinflammatory change was associated with higher IQR/Median ratios, reduced inter-operator concordance and lower ARFI accuracy, as well as a propensity towards increased LSM values (i.e. independent of underlying fibrosis severity).

Our results are in keeping with the body of literature, both for ARFI[16,17,167] and elastography more broadly.[133,134,211] Necroinflammatory change has been consistently associated with higher ARFI velocities in a number of studies,[17,18] which has been widely attributable to the inflammatory infiltrate increasing liver stiffness. Inflammatory activity has also shown to reduce ARFI accuracy,[16,17,167] particularly once the ALT level was over 5 times the upper limit of normal.[17]

Our study findings therefore reiterate the limited role of ARFI in assessing patients with active hepatitis, and caution against the interpretation of elevated LSMs in the presence of necroinflammatory change. Interestingly however, we found that the impact of active

hepatitis on ARFI performance to be considerably weaker than SLD in multi-regression analyses. This further underlies the importance of central adiposity on ARFI performance, elevating SLD above other factors whose impact on ARFI reliability receive arguably greater awareness. This finding is particularly notable, given the significant number of patients with active necroinflammatory change in the biopsy / accuracy sub-cohort.

5.3.2 Hepatosteatosis

The lack of a direct association between hepatosteatosis and ARFI performance was, however, more unexpected. A link was firstly suspected on an anecdotal level, with the department widely perceiving ARFI performance to deteriorate amongst patients with steatotic change. More importantly, however, we expected a relationship would exist on a conceptual level. We hypothesized hepatosteatosis would attenuate the ARFI push pulse and tracking beam, given its well-recognized impact on B mode imaging.[30] This expectation was further heightened by the marked impact that subcutaneous adipose tissue had on ARFI reliability. It was therefore felt to be a natural extension that intrahepatic fat would have similar effects.

The negative finding for hepatosteatosis is therefore difficult to explain. Whilst a type 2 statistical error cannot be completely excluded, it is unlikely that a clinical significant association between hepatosteatosis and ARFI performance would be missed in a cohort of this size. Hepatosteatosis has also failed to be implicated in ARFI performance in what is now a significant body of literature. Therefore our results likely reflect a true negative association between hepatosteatosis and ARFI performance, even though this is conceptually unexpected.

Our results therefore suggest ARFI may still be useful in patients with hepatosteatosis and NAFLD, provided their BMI and SLD are not significantly elevated. In practice these factors almost invariably co-exist with obesity, and therefore elucidating the exact cause of ARFI's poor reliability in NAFLD and the metabolic syndrome may not be clinically relevant in the majority of cases. In Asian populations, however, NAFLD and hepatosteatosis frequently develop at normal or only mildly elevated BMIs.[212] Our results would therefore suggest ARFI may still retain high local performance in the setting of NAFLD and hepatosteatosis, amongst specific patient populations.

Whilst hepatosteatosis did not show a direct impact on ARFI reliability in our study, we found steatosis to have a negative association with absolute LSM (i.e. independent on underlying fibrosis severity). Whilst there is some evidence that hepatosteatosis causes a reduction in LSM values,[32,148] this relationship has not been universally observed in all studies[18,27,31] and is also in conflict with animal experiments.[173,174] Our results therefore provide further evidence that hepatosteatosis causes 'softening' of the liver, thereby resulting in lower ARFI velocities.

5.3.3 Other patient factors

All other patient factors assessed were found not to have an independent impact on ARFI reliability. Whilst increasing age, a diagnosis of NAFLD and hepatosteatosis all appeared to have strong relationships with the IQR/Median ratio, inter-operator concordance and accuracy, these associations were found not to be independent of BMI and SLD in multi-regression analyses. These factors are therefore unlikely to directly impact on ARFI performance, however do so indirectly through their own association with body habitus.

The findings relating to patient age and NAFLD are supported by the majority of ARFI literature. Age has been shown to be not directly linked with ARFI performance in a number of published studies.[94,95] Similarly, a recent meta-analysis has shown ARFI to have good accuracy in quantifying liver fibrosis in the setting of NAFLD.[152]

5.4 Scan factors affecting ARFI reliability

5.4.1 Measurement Depth

We observed ARFI reliability to be reduced amongst measurements acquired within 2cm of the liver capsule. These subcapsular measurements showed both greater deviation from the set's overall median ARFI velocity (i.e. contributing to higher IQRs), but also greater deviation from the liver biopsy result (i.e. indicating lower measurement accuracy). Reliability was particularly reduced when the ROI was positioned within 1.5cm of Glisson's Capsule, however the reliability of measurements performed between 1.5 – 2cm also appeared affected (Table 4.20, Figure 4.7). At the other end of the depth range, ARFI

performance also appeared to slightly reduce when measurements were acquired over 4.5cm deep to the capsule. The impact of deep measurements was however comparatively minor, and was considered to be of insufficient magnitude to warrant further discussion.

The impact of measurement depth on ARFI performance has been previously evaluated in two studies involving tissue phantoms and patients. Sporea *et al.* acquired ARFI measurements amongst 114 patients at three different depths below the liver capsule (0-1cm, 1-2cm and 2-3cm) and found ARFI to have a poorer correlation with histopathology when measurements were acquired within 1cm of the liver capsule (AUROC = 0.469, AUROC = 0.675 and AUROC = 0.714 at the three depths, respectively).[22] Chang *et al.* performed similar analyses amongst a tissue phantom and found lowest measurement variability (i.e. lowest SD) amongst measurements performed at 4 – 5cm depth using the convex probe.[21] On the basis of these combined findings, the EFSUMB recommends ARFI measurements be performed between 1–3cm from the liver capsule, whilst WFUMB guidelines recommend measurements be performed >1.5-2cm from the capsule. Our study findings suggest that ARFI reliability is reduced when measurements are taken within 2cm of the capsule. Whilst this impact was greatest in the immediate subcapsular region (<1cm), measurement reliability was also reduced in the 1 – 2cm bracket. Our results therefore that suggest adopting a more stringent and conservative measurement depth recommendation of >2cm from the liver capsule may be prudent.

The finding of reduced ARFI reliability amongst subcapsular measurements has been most commonly attributed to a band of physiologic fibrosis which normally underlies the liver capsule.[213] This is theorized to result in reduced elasticity in the subcapsular liver, which is therefore non-representative of overall liver stiffness. Our own experience with 2D-SWE, however, suggests that artefact from capsule reverberation is likely to be an equally important contributory factor. The basis for this hypothesis is discussed further in Chapter 7.2.2 below.

Our measurement depth findings do have limitations. Firstly, the small number of patients with contemporaneous liver biopsy results limited our assessments regarding the impact of ROI depth on ARFI accuracy. The analyses are also based on observational data, and therefore do not control for potential confounding variables. The most pertinent factor is operator experience, with inexperienced operators being more likely to acquire subcapsular measurements and to have lower overall ARFI reliability. Sub-analyses were

however performed amongst only experienced operators (i.e. >100 scans performed), and the finding of reduced ARFI reliability amongst subcapsular measurements was maintained, including in the 1 – 2cm depth bracket.

5.4.2 Operator and Institutional Experience

There is very limited existing data looking at the impact of operator experience and training on ARFI reliability. To our knowledge, only two small studies have been published on this subject; which have opposing conclusions. Boursier *et al.* analysed ARFI measurement reproducibility amongst 101 patients scanned by both a novice and expert operator. They found no difference in performance between the two operators and therefore concluded that there is no training effect for ARFI.[34] Ferraioli *et al.* scanned 97 healthy volunteers with a novice and an expert operator, and repeated this over two separate training periods.[33] They found ARFI performance increased during the second scanning session and therefore concluded that ARFI is dependent on operator training.

Our results fall in between these studies, demonstrating a statistically significant, albeit weak relationship between operator experience and ARFI performance. Individual operators who had performed less than 25 ARFI scans showed a trend towards slightly higher IQR/Median ratios (0.173 vs. 0.165, p=0.15), lower inter-operator concordance rates (14.3% vs 11.00%, p=0.002) and lower accuracy (median deviation from biopsy reference range = 0.588 vs. 0.301m/s, p=0.022) than more experienced operators. A similarly minor reduction in ARFI reliability was also observed during the first 150 scans performed in the institution.

Our results build upon the limited existing data on a number of grounds. It is the first to assess the improvement in ARFI performance across the range of operator experience, rather than at two arbitrary time points as with Ferraioli's study. It also assesses the overall training effect for a large number of operators, instead of one or two operators artificially selected for research purposes. We furthermore assessed the impact of operator training on a wide range of performance measures, rather than limiting our analyses to measurement reproducibility. Our results therefore provide arguably the best indication regarding the impact of operator training on ARFI performance to date, and may therefore help inform practice guidelines.

There are no formal recommendations regarding minimum training requirements for ARFI in current elastography guidelines, due to the limited information available for the technique.[35,93] Our results suggest ARFI performance is slightly lower amongst the first 25 scans performed by an operator, and during the first 150 scans performed in an institution. Whether these findings should translate into minimum operator training recommendations for ARFI is however debatable. Some principles can be drawn from the results and established practices for TE. Transient elastography guidelines currently recommend 100 scans be performed by an operator as the minimum training requirement, however 500 scans are required for one to be considered an 'expert'.[35] These recommendations are based on a number of studies,[12,214–216] but most notably the large prospective study by Castera *et al*. which analysed TE failure rates amongst 13,369 examinations.[12] They found operators with less than 500 scans experience to have considerately higher failure on IQR/Median and SR criteria (OR = 2.6, 95%CI: 1.8 – 3.9), which prompted the aforementioned minimum training recommendations. To contrast this with our own results, we found inexperienced operators (≤25 scans performed) to have an OR of 1.2 for failing IQR/Median criteria. The impact of operator training is also negligible when compared to the other reliability factors, with a SLD>2.5cm having an OR of 4.2 of failing the IQR/Median criteria. Therefore whilst ARFI is likely to have a weak training curve, its impact on ARFI performance appears relatively minor and may not justify formal recommendations regarding minimum operator training.

Reason for weak training curve

The limited training curve for ARFI compared to TE likely reflects the lower operator dependence of the point SWE technique. Unlike TE, ARFI utilises a standardised ultrasound push pulse to achieve tissue excitation, which is automatically generated by the ultrasound device. This removes an operator dependent step from the acquisition process, which the founders of the technique originally theorized would result in lower operator variability and dependence.[104] A second possible contributory factor is that ARFI was locally performed by sonographers and radiologist with prior training in ultrasound, rather than hepatologists with no formal ultrasound experience.

There are nonetheless additional operator dependent acquisition factors, external to tissue excitation, which likely contribute to the weak training effect observed in our cohort. Some of these factors include the use of an optimal ultrasound angle (i.e. ensuring the

beam remains perpendicular to the liver capsule to minimize refraction), suitable ROI positioning and the appropriate timing of ARFI acquisitions in relation to breath holds. ROI depth within the liver provides an excellent example of these operator dependent factors; with a higher rate of subcapsular measurements observed amongst operators who had completed ≤25 vs. >25 ARFI scans (2.16% vs 0.76%, p=0.0001). Our findings nonetheless demonstrate that the impact of these additional operator dependent factors is likely to be minor.

Limitations

Our analyses looking at the impact of operator and institutional experience on ARFI performance do, however, have limitations. At the individual operator level, our analyses included a wide variety of operators ranging from experienced radiologists and sonographers through to medical imaging trainees. Junior staff often performed a lower number of ARFI scans compared to more experienced colleagues, reflecting their often transient roles in the department. As a result, the inexperienced operator group (i.e. < 25 ARFI scans performed) are likely to encompass a higher proportion of junior and trainee operators compared to those with greater ARFI experience (i.e. >25 ARFI scans). To help address this bias, sub-analyses were performed amongst only operators who ultimately completed >100 ARFI scans. The results of these sub-analyses confirmed our original findings, and again showed only a weak relationship between scan experience and ARFI reliability (Table 4.23). In any case, the inclusion of transient and junior staff in our analyses would likely result in the overestimation of any training curve present (i.e. by causing an apparent reduction in ARFI performance during lower number of scans performed). This provides further reassurance to the minor impact of operator experience on ARFI performance.

A second limitation is that operators included in the analyses had predominantly low levels (<200 scans) of ARFI experience. Our analyses of training effect are therefore limited to first 200 scans performed by an operator, and we cannot conclude whether ARFI performance gradually improves until a total of 500 scans have been completed; as is the case with TE.[12] Drawing from TE experience, however, the training effect for Fibroscan® was steepest during the first 100 scans performed by an operator. Given that we observed only minor changes in ARFI performance during what should theoretically be

the steepest part of the ARFI training curve, it is likely that any change in performance beyond the completion of 200 ARFI scans is likely to be negligible.

Our study did not analyse the impact of general ultrasound experience on ARFI performance. Having greater conventional ultrasound experience could impact on the ability to obtain an appropriate acoustic window and maintain the correct ultrasound angle. General ultrasound aptitude could therefore theoretically impact on ARFI acquisition, and warrants analysis in future studies.

Finally, our institutional level findings reflect a single centre experience, and may not be directly applicable or transferrable to all centres. Differences in the types of operators and their levels of prior ultrasound and elastography experience could potentially impact on the training curve for a particular institution. Our results also reflect an experience of ARFI introduction during the technology's infancy, when formal acquisition guidelines were still in the process of development. Now that acquisition recommendations have been refined, it may be expected that less evolution in acquisition technique (and therefore less change in ARFI performance) will be observed over time. Our results may therefore overestimate the impact of institutional experience on ARFI reliability in a more contemporaneous setting.

5.5 Strengths and limitations of the ARFI analyses

Patient cohort

Our study adopted deliberately broad inclusion criteria which encompassed all patients with diffuse CLD who underwent an ARFI scan for the clinical assessment of liver fibrosis. This was necessitated by some analyses which required an all inclusive cohort to avoid bias; an example being the impact of operator and institutional experience on ARFI performance. The limited selection criteria resulted in the patient cohort being directly reflective of the local target population and also made our study findings (e.g. SLD and inter-operator concordance) relevant to a broader group of patients.

Chronic liver disease is not a uniform diagnosis, however, and encompasses a highly heterogeneous group of patients with differing aetiologies and disease processes. Drawing conclusions from a highly heterogeneous cohort conversely reduces the direct

applicability of our findings to specific patient groups such as HBV, HCV or NAFLD. Whilst sub-analyses suggest our major findings, including SLD, apply equally to the major CLD sub-types, further validation of these findings may be required in disease specific populations.

There are also a number of differences between our patient cohort and ARFI set-up from those overseas, which may reduce the generalizability of our study findings to other populations. As previously mentioned, the most important difference is likely to be the high obesity rate in our cohort; with approximately 26% of the patient cohort being obese (BMI >30kg/m^2) and a further 32% overweight (BMI 25 – 30kg/m^2). The mean BMI of 27.3 is somewhat higher than that reported in most seminal ARFI papers which generally report a mean BMI of 25 – 26kg/m^2.[150,151,172,217] This difference is likely to be clinically significant, given obesity marked impact on ARFI reliability.

Furthermore, ARFI readings were predominantly taken by sonographers in our study; which is inline with standard operating ultrasound procedure in Australia. This practice differs from overseas centres, however, where ARFI is routinely acquired by a doctor (most commonly a hepatologist). Whether this has any bearing on ARFI performance, positive or negative, has not been assessed in the literature. It does however introduce an additional unknown which could impact on the generalizability of our results to overseas centres.

Strengths and limitations

As has been heavily discussed in prior chapters, the predominant limitation of our ARFI results is the small number of patients that had undergone liver biopsy in the cohort. This not only reduced power, but also introduced issues of selection bias which together limit the findings drawn from our accuracy analyses.

Our study did however have strengths. Firstly it represents one of the largest patient cohorts scanned with ARFI in the literature. It also analysed wide-ranging performance parameters including IQR/Median, inter-operator agreement and accuracy and therefore provides a very broad assessment of ARFI performance.

A particular point of difference is the large number of patients scanned with multiple operators in our cohort. Prior studies have assessed the intra-class correlation co-efficient

(ICC) between two operators, who are predominantly experienced clinicians. These confirm that high inter-operator reproducibility can be achieved with ARFI, however it is difficult to know whether these results are relevant to operators more broadly, particularly those with lower levels of ARFI experience. Our results are therefore valuable in providing a more 'real world' representation of inter-operator agreement amongst operators of varying clinical backgrounds, competencies and experience levels with ARFI. The large number of patients scanned also provides one of the most powerful indications on what influences inter-operator agreement in ARFI. Whilst the use of multiple different operator pairs prevented a formal ICC being calculated for the overall cohort, the very large number of patients assessed negates the influence of variable operator performance on overall operator agreement.

Inter-operator concordance definition

Discussion is also required regarding the definition of inter-operator concordance adopted in this study, with operators considered concordant if they obtained ARFI velocities within the same or adjacent F scores. Unfortunately, there is no precedence as to what constitutes operator concordance in the ARFI literature and the above definition was chosen for a number of reasons.

Firstly, converting a continuous scale (i.e. ARFI velocity) into a discrete variable (i.e. F scores) is problematic, as operators with very consistent ARFI readings may still be ascribed different F scores if they fall on either side of an F score cut-off. As a result, a one score difference in operator F scores may be attributable to this occurrence rather than reflecting true issues with ARFI performance. In contrast, a two or more F score difference is always reflective of true operator disagreement and therefore provides a more consistent indication of ARFI reliability issues.

A more stringent definition of inter-operator discordance was also required in our study, as we planned to trial inter-operator concordance as a potential reliability indicator. Using a broader definition of inter-operator discordance (i.e. any F score difference) would risk labelling a large number of patients as having 'unreliable' measurements; including some patients in whom operators may have obtained very similar ARFI velocities. Hence a two or more difference in F scores was felt appropriate for this application.

Some inferences can also made from studies looking at concordance between LSM values and histopathologic F scores. Bota *et al*. defined discordance between ARFI LSM and biopsy to be clinically significant when a difference in two or more F scores was observed, with Attia *et al*. adopting the same approach in their study.[28,175] These practices are also standard in the Transient Elastography literature.[126] Whilst our study is looking at inter-operator concordance rather than concordance with biopsy, the underlying rationale and principles are nonetheless similar and provide further justification for the definition adopted.

Despite the above justification, the lack of an established definition for inter-operator concordance makes it a less validated and therefore potentially less robust measure of inter-operator agreement. As a consequence, percentage inter-operator deviation is included as the primary measure of inter-operator agreement throughout the results section. Inter-operator concordance was still included in the thesis, however, as it is felt to be a more tangible measure of clinically significant operator disagreement, which make it a more accessible and potentially more clinically useful measure than percentage deviation. It also functions as a bridge to Chapter 5.6.3, in which inter-operator discordance is used as a surrogate marker of ARFI reliability. Hence the inclusion of inter-operator concordance provides a prelude to these analyses, indicating which patients are likely to 'fail' the proposed reliability assessment approach.

5.6 Strategies for predicting ARFI reliability

The major limitation of point SWE compared to more sophisticated elastography techniques (i.e. 2D-SWE and MRE), is its limited facility to assess the validity of obtained shear wave measurements. LSMs are provided with almost no indication regarding the quality of the shear wave propagation being measured, and clinicians are therefore forced to blindly trust the ARFI result with minimal supportive information. Whilst the IQR/Median ratio provides some indication regarding ARFI reliability, this approach is unfortunately imperfect and further strategies are therefore required to allow clinicians to better assess the reliability of acquired LSMs.

5.6.1 IQR/Median criteria

Our results further support the utility of IQR/Median as a reliability metric for ARFI. In our cohort, scans which met the IQR/Median criteria showing significantly higher correlation with liver biopsy than those failing the criteria (rho = 0.291 vs. -0.146). Our results are therefore in keeping with those of Bota *et al.* and Goertz *et al.,* who found higher IQR/Median and SD/Mean ratios to be associated with lower ARFI accuracy, respectively.[31,36]

IQR/Median ratio is unfortunately, however, an imperfect approach. The criteria firstly failed to identify a significant proportion of patients with inaccurate ARFI measurements, with the criteria having a modest sensitivity of 39% in detecting scans with clinically significant discordance between ARFI velocity and histopathology (i.e. ≥2 F score difference). Conversely, five of the 21 patients (i.e. 24%) who failed the IQR/Median criteria actually had accurate ARFI results (i.e. within the same F score).

Our results illustrate the limitations of IQR/Median ratio as a reliability indicator and caution against placing undue reliance on the criteria. There are however no clinically validated alternatives and therefore the ongoing use of the criteria may still remain justified given its modest predictive utility. Our results do however illustrate the importance of developing alternative reliability assessment strategies to augment the existing approach.

5.6.2 Skin-to-Liver Capsule Distance (SLD)

Rationale

We found obesity to be the primary determinant of ARFI performance, and it would therefore seem prudent that body habitus be considered in any assessment of ARFI reliability. Whilst both BMI and SLD would both function as possible reliability indicators, SLD appeared especially suited to this function on a number of levels.

We firstly found SLD to be more strongly associated with IQR/Median and inter-operator concordance than was BMI. Whilst the small size of the liver biopsy sub-cohort prevented any direct comparison of the relative relationship of BMI versus SLD with accuracy, it is nonetheless likely that SLD is equivalent, if not superior to BMI in predicting ARFI

accuracy. SLD also appears to be a better choice from a conceptual level, due its more direct relationship with ARFI mechanics and ultrasound beam attenuation.

SLD also has a number of practical advantages over BMI as a potential reliability indicator. SLD can firstly be measured by a clinician remote from the patient, and can be ascertained in retrospect after the patient has left the department or from historical exams. It requires no patient involvement and therefore avoids the potential embarrassment associated with other anthropological measurements including BMI. SLD values are very quick to obtain, with measurements taking seconds to perform. This time burden is further minimized by the finding that a single SLD measurement is likely sufficient for clinical purposes. Measuring the SLD from a single ARFI screenshot yielded an estimate within 0.07cm of the patient's mean SLD value of ten measurements. This level of error is very small, and unlikely to be of any clinical significance when compared to the large range in SLD values observed between patients (range = 0.96 – 5.50cm).

SLD also appears to be a robust measurement, which is relatively independent of ARFI operator technique. We observed minimal difference in the SLD values obtained between operators, with operator SLD values deviating by a median of 0.13cm (i.e. 5.99%) in each patient. This variation is small and indicates that SLD is primarily a patient dependent factor which is minimally influenced by variables such as probe pressure, ultrasound angle, or acoustic window. SLD is furthermore a highly objective and quantifiable metric which has the capacity to be applied as either a continuous or binary variable; the advantages of which are discussed further below.

It should also be clarified that SLD is retrospectively measured from ARFI screenshots following the completion of ARFI measurements. In contrast to TE, SLD is not assessed prior to ARFI acquisition and therefore cannot be used to inform whether ARFI measurements should or should not be acquired. Instead, it represents an additional parameter which may help clinicians to retrospectively assess whether obtained ARFI LSMs are likely to be valid.

SLD cut-offs

ARFI performance progressively deteriorated with increasing SLD, and this relationship appeared to hold across the full spectrum of SLD depths. Therefore whilst ARFI accuracy

started declining once SLDs exceeded 2.5cm, patients with a SLD >3.0cm appeared to perform exponentially worse than those with an SLD of 2.5 – 3cm. SLD is therefore most powerful as a continuous variable, and should ideally be utilised and interpreted using a non-binary approach.

With this acknowledged, the dichotomization of continuous variables also has some utility and benefit in clinical practice. It provides a rough cut-off to help inform clinicians as to what constitutes a 'reliable' vs. 'unreliable' measurement, which may be difficult to ascertain from a continuous scale. This rationale has been used to recommend an IQR/Median cut-off of 0.30, despite the relationship between IQR/Median and ARFI accuracy also being a continuous one.[129]

Proposing a potential SLD reliability cut-off for clinical practice is therefore somewhat arbitrary. We found ARFI accuracy to start deteriorating once the SLD exceeded 2.5cm (Figure 4.5, Table 4.18), and a reliability cut-off of ≤2.5cm may therefore be clinically appropriate. Patients with a SLD >2.5cm showed significantly higher IQR/Median ratios (median = 0.363 vs. 0.187, $p<0.001$), greater percentage deviation between operators (29.8% vs. 15.9%, $p<0.001$) and lower correlation between ARFI velocity and histopathology (rho = 0.516 vs. -0.188) than those with a SLD ≤2.5cm. This cut-off is however weighted towards sensitivity (i.e. identifying all unreliable ARFI scans), and as a downside identifies a large proportion (22.6%) of patients as having potentially 'unreliable' ARFI measurements. This is a larger proportion of patients than is identified with IQR/Median criteria (15.9% in our cohort), and may be unacceptably high for clinical purposes.

A less conservative cut-off with greater weighting of specificity may therefore be equally appropriate, depending on the clinical setting. A SLD >3.0 is associated with even higher rates of unreliable measurements, showing yet higher IQR/Median ratios (median = 0.409), similar inter-operator deviation (30.5%) and poorer concordance with biopsy (rho = -0.740) when compared to those patients with a SLD of 2.5 to 3.0cm. The SLD >3.0cm cut-off therefore identifies a small group (i.e. 8.6%) of patients who are at particularly high risk of unreliable ARFI readings.

It may therefore be useful to adopt dual SLD cut-offs in clinical practice. In summary, ARFI measurements with an SLD ≤2.5cm are likely to be reliable, those with an SLD

>3.0cm are likely unreliable, whilst measurements with an SLD of 2.5 – 3.0cm fall within a grey zone and should be interpreted with some caution.

Incorporation with IQR/Median ratio

We found SLD to have a much stronger correlation with ARFI accuracy than IQR/Median ratio in multi-regression analyses. This suggests SLD is likely superior to existing criteria in predicting ARFI reliability, and provides further evidence of SLD's potential clinical utility.

Whilst SLD showed a stronger correlation than IQR/Median, both strategies nonetheless showed independent positive correlations with ARFI accuracy in multi-regression analyses. This suggests that the both approaches likely provide incremental information regarding ARFI reliability, which would make sense conceptually given the disparate nature of the parameters being measured. The potential synergism of both strategies is further illustrated in Table 4.28 and Figure 4.8, in which combining SLD (adopting a cut-off of ≤2.5cm) to existing IQR/Median criteria showed superior stratification of ARFI accuracy than either approach alone. We found very high accuracy amongst patients who passed both criteria, with 84.7% of such patients having an ARFI result which was concordant with histopathology. It is therefore likely clinicians can have high confidence in relying on ARFI results amongst patients with an IQR/Median ≤0.30 and an SLD ≤2.5cm; which encompasses over 70% of patients from the greater study cohort.

Our results suggests that routinely using both reliability criteria in consort may allow clinicians to better gauge the reliability of ARFI results. Both indicators are very easy and quick to obtain, and therefore it would likely be practical and logical to routinely assess and co-report both variables in clinical practice.

Areas for future study

Whilst our results provide evidence for the potential utility of SLD as a reliability indicator, further validation is required before the marker can be recommended for routine clinical use. Firstly, the small number of patients who completed a contemporaneous liver biopsy limited the power of our accuracy analyses. The precise magnitude of SLD's impact on ARFI accuracy therefore remains unknown, as does the true synergism between SLD and

IQR/Median criteria. Whilst SLD was identified as a possible reliability metric in our own cohort, these findings need to be replicated by a second, independent cohort for validation purposes. This is particularly necessary as SLD was identified and assessed as a possible reliability metric following review of our ARFI results (i.e. rather than being listed as an original study aim).

The reproducibility of SLD measurement also needs to be formally assessed. Specifically, does the person measuring the SLD from ARFI screenshots have any significant impact on the obtained SLD value? On a conceptual level, the intra and inter-operator reliability coefficients for SLD measurement are likely to be very high. The skin surface and liver capsule are usually clearly identifiable on ARFI screenshots, minimizing any subjective difference in measurement targets. And any millimeter differences caused by imprecise caliper positioning is also likely to be inconsequential relative to the large difference in SLD values observed between patients. Therefore whilst the reproducibility of SLD measurement needs to be formally documented, we anticipate the approach will have both high intra and inter-operator reproducibility.

5.6.3 Multiple operators

The second new strategy trialed in the assessment of ARFI reliability involved scanning patients with multiple independent operators. We found significant variability in the operator ARFI LSMs obtained in some patients, with inter-operator discordance being predominantly associated with other indicators of poor ARFI performance, including obesity. We therefore hypothesized that the presence of inter-operator concordance versus discordance may provide an indirect indication of ARFI reliability.

Our results confirm what would be clinically expected, showing inter-operator agreement to be a useful indicator of ARFI accuracy. In patients for whom operators obtained discordant ARFI F scores, ARFI showed a poor correlation with biopsy (rho = 0.010). In contrast, patients with two concordant operators showed a considerably higher correlation between ARFI LSM and histopathology (rho = 0.392). Whilst assessing patients with three operators further increased the predictive power; patients with three-way operator concordance showing yet higher correlation between ARFI and biopsy (rho = 0.571).

In cases of operator discordance, we found the lower of the two operator ARFI velocities to be the more accurate. This suggests that operator discordance is primarily attributable to an operator having a falsely elevated LSM, which is in keeping with our previously discussed accuracy findings (Chapter 6.1.1). This finding is however drawn from only a small number of patients (n=11), and therefore requires further validation in a larger patient cohort.

Clinical application

Using inter-operator concordance to predict ARFI reliability is conceptually unrelated to either SLD or IQR/Median criteria. Therefore, whilst the small size of the accuracy sub-cohort prevented any assessment regarding the synergism of the three strategies, it is anticipated that scanning patients with multiple operators would provide incremental information regarding ARFI reliability.

In contrast to SLD and IQR/Median strategies, however, scanning patients with multiple operators has significant practical implications for both patients and staff. The approach extends a patient's scan by approximately five minutes (or longer if three operators are used), and is also contingent on an additional operator being available at the time of scanning. It is therefore unlikely practical or justifiable to use the strategy routinely for all patients, and the approach may be best reserved and selectively applied in patients in whom the reliability of obtained ARFI measurements is in question.

Selecting patients who are most likely to benefit from the approach is therefore important. The greatest yield is likely to be in those patients with a high risk of inter-operator discordance, and we found SLD, IQR/Median ratio and higher ARFI velocities (i.e. F3 / F4) to be the factors most strongly predictive of inter-operator deviation in multi-regression analyses. Patients failing either IQR/Median criteria or having a SLD >2.5cm showed the highest rate of inter-operator discordance (23.1%). This encompassed approximately 29.9% of the patient cohort, and may therefore represent a 'high risk' population who should be scanned by multiple operators as a priority. Patients who had an IQR/Median ≤0.3, a SLD ≤2.5cm, but an ARFI F score of either F3 or F4 represented an 'intermediate' risk group who had a 12.5% rate or inter-operator discordance. Whilst these patients had an higher rate of inter-operator discordance than the remaining 'low risk' patients (discordance rate = 12.5% vs. 6.2%), their addition to the target group would result in

50.9% of the total cohort being scanned by an additional operator; which may not be practical for the majority of ultrasound centres. Having these multiple tiers of patient risk do, however, allow departments to implement the strategy at a scale which is appropriate to their logistical circumstance.

A second clinical setting which may benefit from the approach includes patients being scanned by inexperienced operators (i.e. ≤25 ARFI scans completed), or during the first 150 scans performed at an institution. Whilst operator experience showed only a modest impact on ARFI performance in our study, the strategy may nonetheless increase the confidence in obtained results and alert to operator performance issues during these probationary periods.

5.7 Cirrhosis assessment

5.7.1 Study findings

We found ARFI LSMs to have a moderate relationship with overall cirrhosis severity. ARFI demonstrated a moderate positive correlation with both MELD Score (rho = 0.342, $p<0.001$) and Child-Pugh Score (rho = 0.363, $p<0.001$), a weak association with a number of blood test parameters associated with hepatic reserve (e.g. albumin, bilirubin, INR, platelets), as well as modest power in predicting the presence or absence of cirrhotic complications.

Our findings therefore confirm that ARFI LSM values increase with the progression from early to late cirrhosis. This finding was largely expected, given that cirrhosis is not a single entity and instead encompasses a wide spectrum of liver fibrosis severities and clinical states. This is reflected in a number of histopathologic grading systems, including Ishak and Laennec, which sub-categorise cirrhosis into multiple tiers of fibrosis severity which have shown to correlate with the development of cirrhotic complications and patient prognoses.[4,218] Our results therefore confirm that a patient's ARFI LSM does provide some indication regarding the severity of their underlying cirrhosis.

Despite this relationship, significant overlap was observed between the ARFI velocities of patients with early versus late cirrhosis, and those with or without cirrhotic complications (Figures 4.9, 4.10 and 4.11 and Table 4.35). This again suggests there are other important patient and scan variables which determine the ARFI LSM value, beyond the severity of the underlying liver fibrosis. The possible mechanism underlying the modest relationship observed between ARFI LSM and cirrhosis is further discussed in Chapter 5.7.5 below.

As a consequence, ARFI LSM values showed only modest accuracy in identifying the presence or absence of cirrhotic complications in our cohort, including encephalopathy (AUROC = 0.602), ascites (AUROC = 0.580), any oesophageal varices (AUROC = 0.687) and moderate to large oesophageal varices (AUROC = 0.725). Therefore whilst ARFI LSM provides some indication of underlying cirrhosis severity, the strength of the relationship appears insufficient for clinical application. Specifically, some prior studies have proposed employing ARFI as a gatekeeper to gastroscopy; using the tool to identify patients with a low likelihood of oesophageal varices in whom endoscopic screening may not be required.[39,40] Whilst we found ARFI to have greatest accuracy for predicting the presence or absence of oesophageal varices, the level of accuracy and sensitivity observed (even when adopting conservative F score cut-offs) appeared insufficient to safely exclude patients from endoscopic surveillance.

5.7.2 Comparison with existing literature

As outlined in Chapter 2.6.4, the literature surrounding ARFI in the assessment of cirrhosis severity has shown variable and at times conflicting results. Some studies have reported very high levels of accuracy, particularly in the assessment of portal hypertension[37,38] and oesophageal varices,[39–41] whilst an equal number have found poor performance in this setting.[43–45,219]

Our results are most in line with the latter. This includes a study from Bota *et al.*, who found the correlation between ARFI LSM and Child-Pugh Score ($r=0.264$, $p<0.001$), MELD score ($r=0.194$, $p=0.005$) and a range of other blood test parameters (bilirubin, albumin, prothrombin time) to be weaker than was observed in our own cohort.[44] Our accuracy in identifying patients with a MELD score of >15 was similar to that reported by Vermehren *et al.* (AUROC = 0.67 vs. 0.69, respectively), as was our accuracy in identifying patients with Child-Pugh B/C cirrhosis (AUROC = 0.71 vs. 0.69).[45]

In assessing for presence or absence of oesophageal varices, our results were in the intermediate range. Whilst we did not observe the very high accuracy levels described in some studies,[39–41] we found ARFI to have greater predictive utility than others.[43,45,219] Specifically, our AUROC in detecting moderate to large oesophageal varices was significantly higher than reported by Vermehren *et al.,* (AUROC = 0.73 vs. 0.58, respectively), whilst Ye *et al*. found no difference in ARFI LSM between patients with or without oesophageal varices, nor a correlation with variceal grade.[45,219]

Our results therefore add to this somewhat conflicting body of evidence, and echo the cautionary findings reported by a number of studies.

5.7.3 Underlying mechanism

There are multiple possibly contributory explanations for the weak relationship observed between ARFI LSM and cirrhosis severity.[220] Firstly, our results suggest cirrhosis severity and the presence of cirrhotic complications is not wholly dependent on liver fibrosis severity. This may be somewhat expected, as MELD and Child-Pugh reflect hepatic decompensation due to portal hypertension and loss of hepatocyte mass, rather than fibrosis levels. Cirrhosis and portal hypertension are also highly complex entities which involve a manifold of pathologic processes including sinusoidal disruption, development of microthrombi as well as the distortion of the hepatic vasculature with associated shunting.[221] These process may not be associated in a change in liver stiffness and therefore may not be reflected in ARFI LSM values.

There may also be some changes in advanced cirrhosis which could paradoxically reduce liver stiffness. Firstly, the development of oesophageal varices and venous collaterals could theoretically divert blood flow around the liver, potentially reducing hepatic congestion and liver stiffness. The reduction in necroinflammatory activity seen in advanced cirrhosis could also theoretically reduced ARFI velocities, given the well-recognised impact of hepatic inflammation on ARFI LSM values.[16,167]

It is also possible that the factors underlying ARFI's modest performance at lower levels of fibrosis quantification in our study may be similarly applicable to the assessment of cirrhosis severity. Given the marked impact of obesity and SLD on ARFI performance

described above, it is likely that these factors may have an equal impact on ARFI velocities; albeit at a higher LSM baseline.

Finally the retrospective design of the cohort analyses also raises bias as a possible contributory factor, which is further discussed in the limitations subsection below.

5.7.4 Study limitations

Our cirrhosis analyses did have limitations, which primarily relate to the study's retrospective design. The presence of cirrhotic complications was assessed retrospectively from medical records, which, as with all retrospective studies, introduces the possibility of bias in both the documentation of cirrhotic complications in clinical notes and their subsequent interpretation by the researcher. This is further compounded by mild ascites and encephalopathy often being subtle clinical findings, which could potentially be missed if not specifically assessed for in clinical practice. This may have contributed to the particularly low accuracy observed with encephalopathy and ascites in our study.

Analyses relating to the MELD score and the prediction of oesophageal varices are however likely to be less subject to reporting bias. MELD score is an objective measure derived from blood tests alone, whilst the presence or absence of oesophageal varices is carefully evaluated and routinely documented in screening gastroscopy reports. These analyses are therefore likely less dependent on clinician assessment and documentation, and are therefore likely to be relatively robust. Despite this, ARFI LSM continued to show only a modest association with both MELD score and the presence/absence of oesophageal varices, which suggests reporting bias is unlikely to fully explain the modest performance observed in our study.

Our study also did not evaluate ARFI splenic stiffness in the assessment of cirrhosis severity, which has been shown to be a more accurate indicator of cirrhosis severity in a number of clinical trials.[219,222] It provides an indirect assessment of splenic congestion, and therefore has shown greatest utility in assessing the presence / severity of portal hypertension and portal hypertensive complications; particularly oesophageal varices.[223,224] It is therefore possible that ARFI may still have high utility in locally assessing cirrhosis severity using this alternative approach.

Chapter 6. Toshiba 2D-SWE Results

6.1 Patient Characteristics

The cohort consisted of 55 patients, of which 55% were male. The age range was 21 to 89 years, with the median age being 52 years. The most common liver disease aetiologies included non-alcoholic fatty liver disease (NAFLD, n=17), hepatitis B (n=16), hepatitis C (n=9) and alcohol (n=7). Fourteen patients in the cohort (25%) had either moderate or severe hepatosteatosis based on B-mode imaging. Patients with a normal BMI (<25kg/m^2, n=20), overweight BMI (25–30kg/m^2, n=19) and obese BMI (>30kg/m^2, n=16) were all represented in the cohort. The median overall SLD of patients was 1.86cm (q1-q3: 1.52-2.14cm). Full cohort characteristics are summarised in Table 6.1.

Table 6.1: Demographics and chronic liver disease characteristics of the patient cohort.

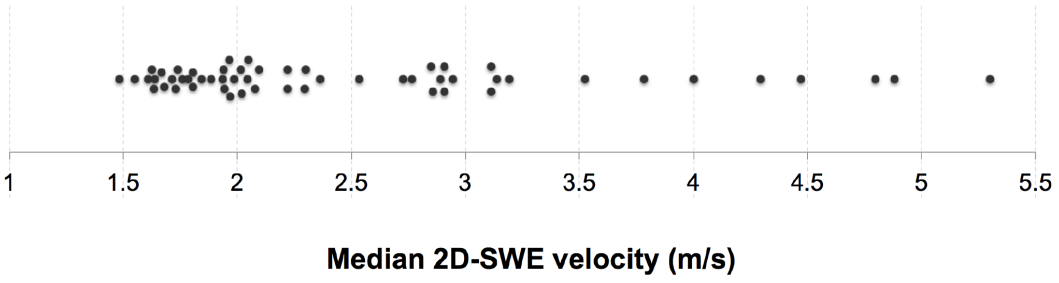
Patient Characteristics	**Number of Patients (% of cohort)**
Gender	
Male	30 (55%)
Female	25 (45%)
Age (years)	
20 - 40	13 (24%)
40 - 60	28 (50%)
60 - 80	13 (24%)
>80	1 (2%)
Liver disease etiology (based on physician assessment)	
NAFLD	17 (31%)
Hepatitis B	16 (29%)
Hepatitis C	9 (16%)
Alcohol	7 (13%)
Cryptogenic	3 (6%)
Drug-induced liver injury	2 (4%)
PBC	2 (4%)
Cardiac cirrhosis	1 (2%)
Autoimmune hepatitis	1 (2%)
Other	4 (7%)

Body Mass Index (BMI, kg/m^2)	
<25	20 (36%)
25 – 30	19 (35%)
30 – 35	8 (14.5%)
>35	8 (14.5%)
Hepatosteatosis (based on B-mode imaging)	
No	26 (47%)
Mild	15 (27%)
Moderate	10 (18%)
Severe	4 (7%)
Skin-to-Liver Capsule Distance (set mean)	
<1.5cm	12 (22%)
1.5 to 2cm	21 (38%)
2 to 2.5cm	14 (25%)
>2.5cm	8 (15%)

6.2 2D-SWE velocities

A large variation in shear wave velocities was observed between individual patients (Figure 6.1), with the range in overall median LSM being 1.49 to 5.30m/s. The median overall 2D-SWE velocity of patients was 2.10m/s (q1-q3: 1.81–2.89m/s).

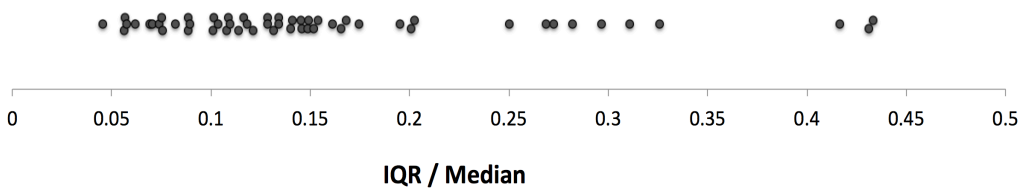
Figure 6.1: Spread in median shear wave velocities of the cohort's 55 patients.



6.3 Measurement Variability

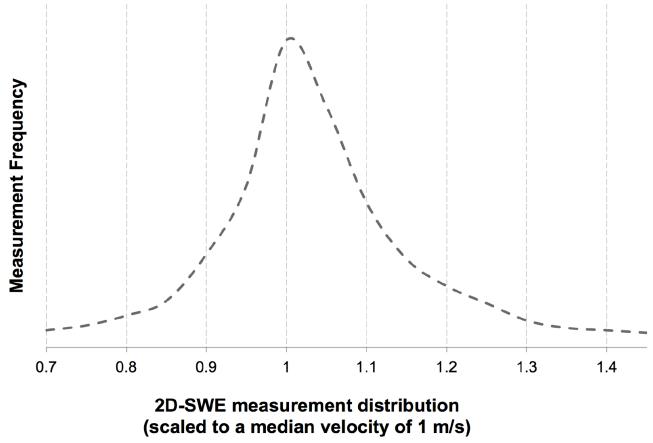
There was a low spread in measurement velocities obtained within each patient, with the patient cohort having a median interquartile range (IQR) of 0.275 (q1-q3: 0.180–0.575) and a median IQR/Median ratio of 0.131 (q1-q3: 0.089–0.174). The spread in IQR/Median values is demonstrated in Figure 6.2, with five patients (9.1%) having an IQR/Median ratio >0.30. The IQR/Median ratio showed a strong correlation with the overall 2D-SWE velocity of individual measurement sets (rho = 0.563, p=0.001). There was very high internal consistency between readings obtained within each patient, as reflected by a high Cronbach's alpha of 0.964 for 10 readings and 0.937 for 5 readings.

Figure 6.2: Spread in IQR/Median values amongst the cohort's 55 patients.



2D-SWE velocities from each of the 55 measurement sets did not follow a normal distribution, with kurtosis ranging between 3.2 and 5.6 and skewness ranging between 1.0 and 1.6. Individual 2D-SWE sets followed a Gamma distribution, with measurements slightly skewed towards higher shear wave velocities (Figure 6.3).

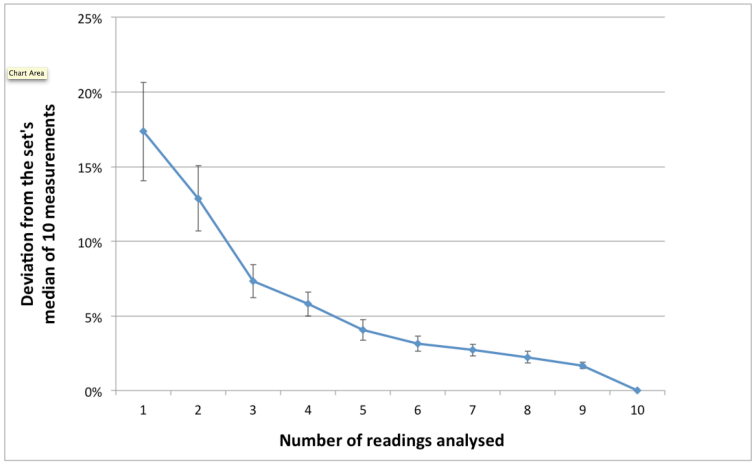
Figure 6.3: Overall distribution of 2D-SWE measurements within each measurement set, scaled to a median of 1. 2D-SWE measurement sets showed a slight positive skew towards higher shear wave velocities.



6.4 Required Measurement Number

The number of measurements required to provide a close approximation to the set's median of ten measurements was analyzed. When increasing numbers of measurements were obtained/analyzed, the calculated velocity became predictably closer to the set's overall median of 10 readings (Figure 6.4). The median of five measurements provided a velocity estimate within 0.11m/s or 4.2% of the set's overall median of 10 measurements. The Bland Altman limits of agreement for five measurements compared to the set's overall median velocity was -0.254 to 0.374m/s.

Figure 6.4: Closeness in approximation to set's overall median of 10 measurements, according to the number of measurements obtained/analysed (mean deviation +/- SEM). Five measurements yielded a liver stiffness approximation below the 5% deviation threshold.



6.5 Factors affecting measurement reproducibility

The association between patient factors and measurement reproducibility (i.e. IQR/Median) was assessed in univariate analyses (Table 6.2). BMI was the primary factor associated with increased IQR/Median ratios (rho=0.388, p=0.01), with overweight and obese patients (BMI >25g/m^2) demonstrating higher IQR/Median ratios than those with normal BMI (median = 0.149 vs. 0.112, Mann Whitney, p=0.011). No significant associations were observed with additional patient factors, however trends towards higher IQR/Median ratios were seen with moderate to severe hepatosteatosis on ultrasound (0.141 vs. 0.122, Mann Whitney, p=0.31), increasing age (rho=0.131, p=0.34) and a clinical diagnosis of NAFLD (0.141 vs. 0.117, p=0.15). Increasing liver fibrosis (as

diagnosed on ARFI) showed no significant correlation with IQR/Median values (rho=0.06, p=0.65).

Table 6.2: IQR/Median ratio amongst different subsets of the patient cohort. BMI and skin-to-liver capsule distance (SLD) showed the strongest associations with IQR/Median ratio.

Patient Characteristics	**IQR/Median Median (q1-q3)**	**Significance of IQR/Median differences (p value) Kruskal Wallis**
Gender		
- Male	0.137 (0.088-0.269)	0.636
- Female	0.129 (0.101-0.161)	
Age (years)		
- <40	0.110 (0.102-0.149)	0.344
- 40-60	0.130 (0.104-0.136)	
- >60	0.145 (0.089-0.282)	
Body Mass Index (BMI, kg/m^2)		
- <25	0.112 (0.085-0.128)	0.013
- 25 – 30	0.134 (0.074-0.296)	
- >30	0.165 (0.138-0.259)	
Liver disease etiology		
- NAFLD	0.141 (0.129-0.250)	0.145
- Hepatitis B	0.115 (0.072-0.154)	0.129
- Hepatitis C	0.118 (0.110-0.269)	0.570
- Alcohol	0.108 (0.088-0.152)	0.579
- Other	0.168 (0.119-0.174)	0.757
Hepatosteatosis		
- No	0.123 (0.089-0.168)	0.543
- Mild	0.121 (0.088-0.161)	
- Moderate/Severe	0.141 (0.108-0.272)	
SLD (Set mean)		
- <1.5cm	0.105 (0.070-0.119)	0.002
- 1.5 to 2cm	0.118 (0.890-0.142)	
- >2cm	0.164 (0.141-0.272)	
F-score (as defined by ARFI)		
- F0/1	0.128 (0.078-0.163)	0.65
- F2	0.137 (0.107-0.163)	
- F3	0.371 (0.262-0.432)	
- F4	0.124 (0.092-0.159)	

Skin-to-Liver Capsule Distance (SLD) also showed a moderately strong correlation with IQR/Median ratios (rho=0.426, p=0.002, Table 6.2). Measurements with an overall SLD >2cm showed considerably greater deviation from the set's median of 10 readings than those with a SLD ≤2cm (mean deviation = 0.501 vs. 0.268m/s, Table 6.3). As seen in the ARFI cohort, SLD again showed a strong correlation with BMI (rho=0.787, p=0.01).

Table 6.3: Deviation of individual 2D-SWE measurements from the set's overall median of 10 readings, according to SLD. Measurements with higher SLDs showed overall greater deviation from the set's median velocity.

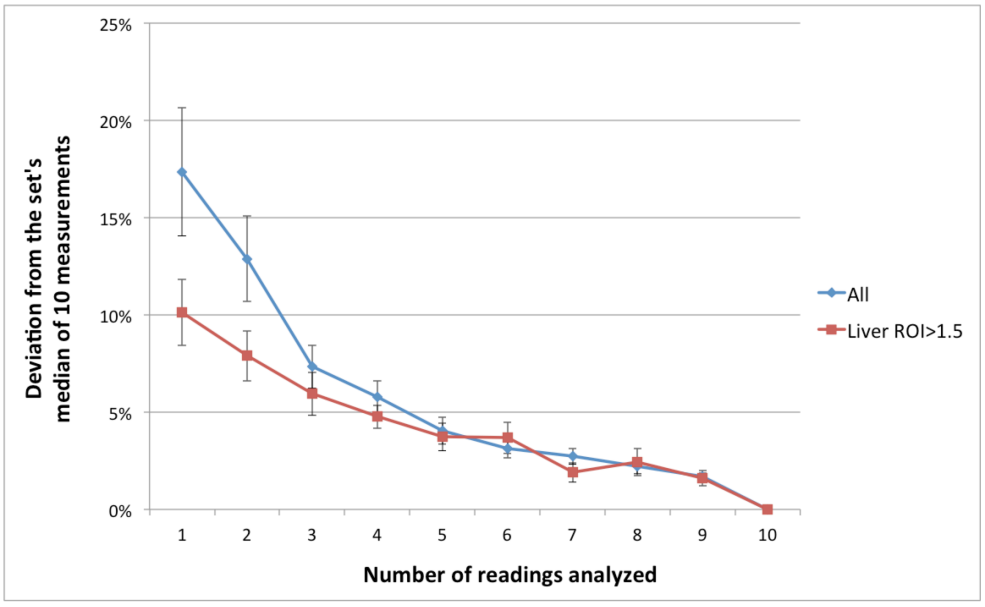
Skin-to-Liver Capsule Distance (SLD)	**Deviation of individual measurements from the set's overall median of 10 readings. (Mean +/- SD)**
<1.5 cm	0.140m/s (+/- 0.15m/s)
1.5 to 1.99cm	0.289 m/s (+/- 0.41m/s)
2 to 2.49cm	0.388 m/s (+/- 0.40m/s)
≥2.5cm	0.463 m/s (+/- 0.77m/s)

A greater spread in LSMs was observed when the centre of the measurement ROI was positioned within 1.5cm of the liver capsule (Table 6.4). These subcapsular measurements showed significantly greater deviation from the set's overall median speed (mean deviation = 0.39m/s) than measurements taken >1.5cm deep to the liver capsule (mean deviation = 0.22m/s, unpaired t test, p<0.001). When measurements taken within 1.5cm of the liver capsule were excluded from analyses, 2D-SWE measurements showed improved overall internal consistency; with only four samples being required to achieve an estimate within 5% of the set's median of 10 readings (Figure 6.5).

Table 6.4: Relationship between measurement depth in the liver (i.e. distance from the liver capsule to ROI centre) and measurement reliability. Measurements obtained within 1.5cm of the liver capsule demonstrated increased deviation from the set's overall median velocity and higher ROI SD / Speed values.

Capsule to ROI Distance	Median deviation from the set's overall velocity (m/s)	Median ROI SD / Speed
<1cm	0.578	0.169
1 to 1.5cm	0.284	0.122
1.5 to 2cm	0.191	0.109
>2cm	0.218	0.119

Figure 6.5: The closeness in approximation to the set's overall median of 10 measurements achieved by acquiring / analysing an increasing numbers of readings. The exclusion of measurements obtained within 1.5cm of the liver capsule (red line) reduced overall measurement variability. This resulted in fewer measurements being required to achieve a reliable estimate of liver fibrosis; the 5% deviation threshold being crossed following the acquisition of four measurements.



6.6 ROI SD / Speed ratio

The ROI SD / Speed ratio reflects the variability in shear wave velocities recorded within the measurement ROI. Measurements with high ROI SD / Speed ratios had relatively poor internal consistency, showing greater mean deviation from the set's overall median (Table 6.5). The ROI SD / Speed showed a moderately strong correlation with the absolute deviation (in m/s) of measurements from the set's overall median (rho=0.330, p=0.001). Measurements with a ROI SD/Speed ratio >0.15 showed greater overall deviation from the set's overall median than those with a ROI SD/Speed ≤0.15 (mean absolute deviation = 0.421m/s vs. 0.219m/s, unpaired t test, p=0.0001).

Table 6.5: Relationship between ROI SD/Speed and the deviation of individual measurements from the set's median velocity.

Individual measurement ROI SD/Speed	Number of readings (%)	Mean absolute deviation from the set's overall median	Mean percent deviation from the set's overall median (%)
<0.05	32 (6%)	0.099 m/s	5.73%
0.05 – 0.099	152 (28%)	0.206 m/s	8.10%
0.10 – 0.149	134 (25%)	0.272 m/s	9.76%
0.15 – 0.249	141 (26%)	0.400 m/s	13.17%
≥0.25	78 (15%)	0.454 m/s	14.70%

Patient factors associated with increased ROI SD/Speed ratios included higher BMI (rho=0.444, p=0.001) and longer SLDs (rho=0.518, p=0.0001). Weaker correlations were also seen with increasing age (rho=0.329, p=0.015) and the clinical diagnosis of NAFLD (p=0.02). The breakdown of ROI SD/Speed values across the patient cohort is listed in Table 6.6.

Subcapsular 2D-SWE measurements also showed higher overall ROI SD/Speed values than those taken more deeply in the liver (Table 6.4). The median ROI SD/Speed value of measurements positioned within or beyond 1cm from the liver capsule was 0.169 vs. 0.117 (Mann Whitney, p<0.001).

Table 6.6: Median ROI SD/Speed values amongst different demographic and clinical subsets of the patient cohort.

Patient Characteristics	**ROI SD / Speed Median (q1 - q3)**	**Significance of difference in ROI SD / Speed (p value)**
Gender		
- Male	0.155 (0.101 - 0.208)	0.171
- Female	0.123 (0.081 - 0.174)	
Age (years)		
- <40	0.104 (0.077 - 0.174)	0.09
- 40-60	0.132 (0.101 - 0.174)	
- >60	0.183 (0.157 - 0.212)	
Body Mass Index (BMI, kg/m^2)		
- <25	0.101 (0.079 - 0.119)	0.003
- 25 – 30	0.153 (0.088 - 0.178)	
- >30	0.180 (0.154 - 0.211)	
Liver disease etiology		
- NAFLD	0.177 (0.151 - 0.196)	0.043
- Hepatitis B	0.113 (0.101 - 0.201)	0.882
- Hepatitis C	0.122 (0.113 - 0.222)	0.946
- Alcohol	0.147 (0.099 - 0.177)	0.840
- Other	0.174 (0.125 - 0.187)	0.105
Hepatosteatosis		
- No	0.123 (0.123 - 0.187)	0.351
- Mild	0.153 (0.153 - 0.184)	
- Moderate/Severe	0.158 (0.158 - 0.195)	
SLD (Set mean)		
- <1.5cm	0.081 (0.073 - 0.107)	<0.001
- 1.5 to 1.99cm	0.147 (0.099 - 0.177)	
- 2 to 2.5cm	0.165 (0.132 - 0.184)	
- >2.5cm	0.185 (0.143 - 0.243)	

Chapter 7. Discussion – 2D-SWE (Toshiba)

7.1 Measurement variability

We found the new 2D-SWE system to have high internal measurement reproducibility in the assessment of liver fibrosis. This was evidenced by both very high Cronbach's alpha values (0.937 and 0.964 for 5 and 10 measurements respectively), but also the system's low median IQR/Median ratio of 0.131; which is lower than that widely reported in both ARFI and Fibroscan®.[36,126,167]

The low measurement variability of 2D-SWE also translated into a small number of measurements being required to provide an adequate estimate of liver stiffness. Acquiring five 2D-SWE measurements showed a Cronbach's alpha value of 0.937, and yielded a liver stiffness estimate within 5% of the overall median velocity of ten readings. A standard error of less than 5% is generally considered acceptable in elastography practice, and this principle has been used to make recommendations regarding required measurement number for both ARFI and SSI.[95,204] Karlas *et al.* performed ARFI within a cohort of 50 healthy individuals, and found eight samples were required to achieve an estimate within 5% of the set's overall median of ten measurements.[95] Yoon *et al*. also applied a similar threshold to recommend six measurements be acquired with SSI to yield a LSM estimate with less than 5% error.[204] Applying the same threshold to our own cohort, acquiring five measurements would appear sufficient for the Toshiba 2D-SWE system. The level of imprecision observed with five samples (mean deviation of 0.11m/s from the median of 10 measurements) is also likely to be clinically negligible, in view of the wide range in LSM values observed between patients (range = 1.49 – 5.30 m/s, Figure 6.1).

The low measurement variability observed for Toshiba 2D-SWE is also encouraging for future accuracy analyses. Measurement variability, as indicated by IQR/Median ratio, has been shown to be a powerful predictor of accuracy for both ARFI[31,36] and Transient Elastography.[126,129] The high measurement consistency observed with Toshiba 2D-SWE will hopefully then translate into high accuracy for the technique; as has been the case for SSI.[24,225]

The low intrinsic measurement variability of the Toshiba 2D-SWE technique is likely attributable to a number of factors. Firstly, the Toshiba 2D-SWE measurement ROI is

larger than ARFI; encompassing an area of 0.79cm^2 compared to 0.50cm^2, respectively. This allows each measurement to sample a larger area of tissue, providing a superior representation of overall liver stiffness per reading. A more important factor is however likely to be the improved visualization of regional shear wave propagation afforded by the Speed 'Smart Map' and 'Propagation Map'. The superior visualization of shear wave characteristics has a number of theoretical advantages, which may improve the reliability and internal consistency of obtained Toshiba 2D-SWE measurements. It firstly enables operators to better assess 'single shot' acquisitions, allowing operators to reject acquisitions which are of insufficient quality for quantitative analysis. The information may also assist in the optimisation of ROI positioning, enabling operators to avoid regions with heterogeneous propagation characteristics or artefact which may yield aberrant results. This qualitative information regarding shear wave propagation is not provided by ARFI, and whilst other 2D-SWE techniques offer an equivalent to the Speed Smart Map, the Propagation Map is unique to the Toshiba 2D-SWE system. The two display modes are purported to provide differing but complementary information regarding shear wave propagation, and we anecdotally observed cases in which aberrations in shear wave propagation were only apparent on the Propagation Map (e.g. Figure 3.8 - 2a and 2b). Whether this theoretical advantage over ARFI and other 2D-SWE techniques actually translates into improved performance, however, requires assessment in future head-to-head clinical trials.

Our results also likely overestimate Toshiba 2D-SWE measurement variability in optimal conditions. Study operators were inexperienced with 2D-SWE at study commencement, and a number of subcapsular measurements were acquired as a result. These measurements are associated with reduced measurement reproducibility (described below) and their inclusion in composite analyses may have increased the observed IQR/Median. Furthermore, our study had a relatively high prevalence of obesity, which may similarly elevate the overall measurement variability observed (also discussed below). The recommendation of a minimum of five measurements is therefore likely to be conservative, and fewer readings may be acceptable under optimal conditions. This is illustrated in Figure 6.5, in which the exclusion of subcapsular measurements alone resulted in the 5% deviation threshold being reached following the acquisition of four measurements. Five readings are however likely to be an appropriate recommendation for 'real-world' conditions and amongst similar patient populations.

We also found that 2D-SWE measurements followed a gamma rather than normal distribution, with measurements skewed towards higher shear wave velocities. In skewed data sets, the median value is considered to be a more statistically robust representation of the set's middle than is the mean. This principle has been previously used to recommend the use of the set's median value for analogous elastography techniques, including ARFI, TE and SSI.[35] Our results suggest that a similar practice should therefore be adopted for Toshiba 2D-SWE.

Comparison with SSI recommendations

Our recommendation of five measurements for Toshiba 2D-SWE is higher than the minimum of three recommended for SSI in the EFSUMB guidelines.[35] This difference is likely to have multiple contributory explanations. Firstly, whilst the EFSUMB recommend three measurements as a minimum requirement, there is significant variability in the SSI literature with individual groups recommending between three and six measurements as optimal.[186,204] This is reflected in the major SSI papers, which routinely acquire either four or more commonly five SSI measurements per patient.[24,176,226,227] Hence the EFSUMB recommendation of 3 measurements represents a minimum standard, which may not be universally supported or adopted by all groups.

The measurement number recommendation may also heavily influenced by research method. The EFSUMB recommendation appears to be drawn from a study by Sporea *et al.*[186] who compared the strength of the correlation between SSI and TE when the mean of three SSI measurements were used compared to mean or median of five. They found that the mean of three SSI measurements provided an equivalent correlation with TE than the mean or median of five SSI measurements ($r = 0.691$ vs. $0.711 / 0.683$). The study didn't compare the strength of correlation against liver biopsy. And more importantly, they used only five SSI measurements as the optimal reference standard (i.e. rather than comparing to a median of ≥10 samples). Given that some groups have found six measurements to be the minimum required to maintain precision, having equivalent precision to 5 measurements may not necessarily guarantee precision has been maintained. Yoon *et al.* used similar principles to our own study, using 5% deviation as the threshold to maintain measurement precision.[204] They concluded that six SSI measurements were required to

maintain these precision targets, and it therefore appears that this research approach may yield a more conservative recommendation.

It is debatable as to which approach is optimal and therefore should be used to guide acquisition guidelines. Ultimately it comes down to achieving an appropriate balance between maintaining measurement precision and the logistical and financial constraints associated with acquiring numerous measurements. Given the uncertainties regarding which approach is optimal, we believe ensuring accuracy should take precedence over the minor time saving associated with acquiring two fewer measurements. Hence we feel adopting a research method which may generate a more conservative measurement number recommendation may not be undesirable.

The difference between our recommendation of five measurements for Toshiba 2D-SWE compared to three for SSI therefore should also not be used to infer the relative measurement variability of the two technologies. There remains no head-to-head assessment of the two technologies in the published literature, and the differences in the research approaches employed, operator characteristics and patient cohorts precludes direct comparison between studies.

7.2 Factors affecting 2D-SWE measurement variability

7.2.1 Body Habitus

Similar to ARFI, body habitus was again found to be the predominant factor influencing measurement variability for the Toshiba 2D-SWE technique. Increased BMI and SLD values were both associated with higher IQR/Median ratios, with particularly poor measurement consistency observed once the BMI exceeded 30 kg/m^2 or the SLD exceeded 2.5cm.

The 2D-SWE results provide yet further evidence to a likely 'class effect' of obesity on ultrasound-based SWE tools in general, adding to existing data for TE,[12,23] ARFI,[23,24] and SSI.[24,183,188] Whilst the impact of obesity on Toshiba 2D-SWE accuracy requires future assessment, our findings provide early caution that the technology's overall performance is likely to be similarly dependent on body habitus.

Similar to the above ARFI analyses, SLD again appeared to be more closely related to 2D-SWE measurement reliability than was BMI; with SLD showing a stronger positive correlation with IQR/Median ratio (0.426 vs. 0.388). This provides further weight to previously discussed hypotheses implicating subcutaneous adipose tissue in the degradation of ultrasound push pulses and tracking beams; technology shared by both the point SWE and 2D-SWE techniques. These hypotheses are also supported by the finding of elevated ROI SD / Speed ratios amongst patients with higher SLD values. This indicates that central adiposity results in more heterogeneous and noisy shear wave profiles; reflecting degradation in the regional shear wave propagation characteristics. This would again support the hypothesis from Palmeri *et al.*, who speculated that increasing attenuation of the ARFI push pulse would result in a lower energy transfer within the region of excitation, and thereby less uniform shear wave generation.[170]

The strong impact of SLD on measurement variability also has potential relevance for our recommendations relating to required measurement number. Whilst five measurements are likely adequate for the majority of patients in clinical practice, patients with high SLD values (i.e. >2.5cm) are likely to have poor consistency between measurements and would therefore require more samples to provide a reliable LSM estimate (i.e. with <5% error). This approach of tailoring measurement number according to patient factors is not employed in current elastography practice; with all clinical guidelines recommending a standard number of measurements be taken routinely for all patients.[35,93] Whilst the approach does increase the complexity of Toshiba 2D-SWE acquisitions, it would allow measurement number to be optimized for the individual patient. This could potentially minimize scanning time in patients who are likely to have low measurement variability (e.g. those with an SLD ≤2.5cm), or conversely increasing the number of samples acquired and thereby improving accuracy in patients who are likely to have poor measurement consistency (i.e. those with a SLD >2.5cm). This approach would however require further assessment and validation, to help define the optimal number of measurements required in different patient groups.

Our findings also provide impetus for the assessment of SLD as a possible reliability metric in 2D-SWE. Given the shared technological principles underpinning point SWE and 2D-SWE, SLD may have similar utility in stratifying the reliability of Toshiba 2D-SWE. This too requires further assessment in a patient cohort with a histopathologic reference, to allow the relationship between SLD and Toshiba 2D-SWE accuracy to be assessed.

7.2.2 Measurement Depth

We found measurement consistency to be poor when the center of the measurement ROI was positioned within 1.5cm of the liver capsule. This finding is largely in keeping with the aforementioned ARFI literature, which have found reliability to be reduced when measurements are acquired within the first 1cm underlying the liver capsule.[21,22] They also mirror our own experience with ARFI, however Toshiba 2D-SWE measurement reproducibility only appeared to be reduced within 1.5cm of the liver capsule, compared to 2.0cm for ARFI. This slight variation is of uncertain significance, however one possible explanation may be differences in statistical power. Namely the 2D-SWE cohort was small compared to ARFI (n = 55 vs. n = 943), and may have had inadequate power to detect the milder reduction in measurement reliability observed in the 1.5 to 2.0cm depth range with ARFI.

Our results are however in opposition to existing measurement depth recommendations for SSI, which currently recommend SSI readings be acquired at a depth of 1 to 2cm from the liver capsule.[35] This recommendation is drawn from the study by Wang *et al.* who assessed the impact of SSI measurement depth in a cohort of 30 healthy volunteers in addition to phantom models.[187] They found that measurements showed lowest variability and highest success rates when acquired 3 – 5cm from the probe surface in the phantom model and 1 – 4cm from the liver capsule in patients; and fused these findings to conclude that measurements should be taken between 1 – 2cm from the liver capsule. The study, however, based their findings on measurement success rate, which was found to be 100% across the 1 – 4cm subcapsular depth bracket. Success rate is an insensitive marker of measurement reliability, and therefore the study was not appropriately designed or powered to detect the milder drop in measurement reliability observed in the 1 – 2cm depth range with ARFI and Toshiba 2D-SWE.

Our results suggest that Toshiba 2D-SWE should not be drawing from the SSI recommendations of Wang *et al,* and that measurements should instead be acquired a minimum of 1.5cm deep to the liver capsule. In view of our ARFI results and the shared technology underpinning the elastography techniques, it could be further argued that adopting a more conservative cut-off (i.e. >2cm) may also be prudent. We did not find any reduction in internal measurement consistency with deep measurements, as has been reported in both ARFI and SSI.[21,187] Of note, there were very few measurements (0.83%)

obtained >3cm deep to the liver capsule, and the reliability of deeper measurements cannot therefore be determined from our results.

As discussed in Chapter 6.4.1, the reduced reliability of subcapsular measurements in SWE has been predominantly attributed to a band of physiologic fibrosis underlying the liver capsule, causing disproportionately high shear wave velocities relative to the remaining liver parenchyma.[213] Our experience with 2D-SWE suggests subcapsular reverberation artefact is likely to be an equally important contributory factor. This is firstly suggested by 2D-SWE Speed Smart Maps, which frequently show a band of heterogeneous shear wave characteristics immediately underlying the liver capsule (Figure 3.8 - 2A). This regional degradation of shear wave propagation characteristics is further evidenced by the finding of higher ROI SD/Speed ratios amongst measurements taken within 1.0cm of liver capsule (median ROI SD/Speed = 0.169 vs. 0.002). Finally superficial measurements have also shown reduced performance in phantom models, in which the elasticity characteristics of the medium should be uniform throughout.[21,187] We therefore hypothesize that the low frequency and high absorption characteristics of the ultrasound push pulse may result in local energy transfer as it traverses the liver capsule interface. This would theoretically result in capsular reverberation and thereby the local distortion of local shear wave propagation characteristics.

7.3 ROI SD / Speed

The ROI SD/speed ratio reflects the variation in shear wave velocities recorded within the measurement ROI on the Speed Smart Map. It provides a quantifiable indication of the homogeneity of the shear wave profile within the measurement ROI; low ROI SD/Speed ratios reflecting measurements with uniform shear wave velocities, whilst noisy or heterogeneous shear wave profiles translate into higher values.

We hypothesized that more heterogeneous shear wave characteristics (i.e. higher ROI SD/Speed values) would result in less reliable 2D-SWE measurements, and our results support this theory. We found ROI SD/Speed to show a strong positive relationship with the amount individual measurements deviated from the set's median, with internal measurement consistency becoming particularly poor once the ROI SD/Speed exceeded 0.15.

Our results suggest ROI SD/Speed may therefore have utility as an objective and quantifiable indicator of measurement reliability. The underlying principle is very similar to the IQR/Median ratio, but applied at the level of the individual measurement. Rather than looking at the variability in velocities between measurements to indicate the reliability of the overall measurement set (as for IQR/Median), ROI SD/Speed looks at the variability of shear wave velocities within the measurement ROI to indicate the reliability of that individual measurement.

ROI SD/Speed has the advantage of being readily assessable during scan acquisition and therefore has real potential for clinical application. ROI SD/Speed could potentially assist in the optimization of ROI positioning, providing operators with an additional quantifiable indicator of shear wave uniformity in addition to the Speed Smart Map and Propagation Map. The value could potentially alert operators to issues in acquisition technique and may present a means of providing dynamic feedback during operator training or technique modification. Perhaps the most important potential application of ROI SD/Speed, however, would be in stratifying the reliability of individual measurements obtained within each patient. This could help clinicians determine which measurements are likely to be most accurate in patients where highly variable shear wave measurements are obtained (i.e. high IQR/Median ratios). It could also be used to reject individual readings with high ROI SD/Speed values from the ultimate measurement set. Or alternatively could signal the need to acquire more measurements in a patient, if readings showed persistently high ROI SD/Speed values.

Further analyses looking at the relationship between ROI SD/Speed and 2D-SWE measurement accuracy are however required, as this will ultimately determine the true value of ROI SD/Speed as a possible reliability metric. Future studies are also required to further explore how ROI SD/Speed can be most effectively applied in clinical practice, which may involve some of the hypothesized examples briefly outlined above.

Our study did not assess whether the ROI SD or the ratio of ROI SD/Speed provided a more powerful indication of individual measurement reliability. We chose to use the latter in our study, as we aimed to replicate the concept of IQR/Median but at an individual measurement level. We also hypothesized ROI SD may be more dependent on the liver stiffness of a patient, with high values anticipated to be seen in patients with advanced fibrosis as well as those with heterogeneous shear wave characteristics. We hypothesized

that using the ratio of ROI SD/Speed would help counteract some of this effect, potentially making ROI SD/Speed more closely reflective of shear wave uniformity. This mirrors the use of IQR/Median ratio rather than IQR in assessing the reliability of overall measurement sets. Our results do not provide any confirmation, however, as to whether ROI SD/Speed is superior to ROI SD in practice. This would again be most appropriately assessed in a cohort with a histopathologic correlate, to determine which measure is more closely associated with 2D-SWE accuracy.

ROI SD/Speed may hold relevance to other 2D-SWE techniques. Devices including SSI and the Philips Epiq platforms also provide a standard deviation of shear wave readings within the measurement ROI and therefore a similar approach could be applied to these technologies. As discussed in Chapter 2, a number of technological differences do however exist between the 2D-SWE systems which render their LSM non-equivalent. This is exemplified in the unit differences between the systems, with SSI providing LSMs in kPa and Toshiba 2D-SWE in m/s. As a result, ROI SD/Speed (or the equivalent in kPa) would need to be independently validated as a reliability metric for each 2D-SWE platform. What constitutes an unreliable measurement would also need to be defined for individual devices, as the technical differences prevent the transfer of our own cut-off (ROI SD/Speed>0.15) between systems. There is some early evidence supporting the use of such an approach for SSI. Thiele *et al.* showed that SSI measurements with a lower standard deviation of shear wave velocities within the measurement ROI (ROI SD ≤1.75kPa) showed higher accuracy.[228] The ratio of ROI SD/Speed has not however been assessed previously to our knowledge. Given its possible theoretical advantages over ROI SD, exploration of ROI SD/Speed as a potential reliability indicator for 2D-SWE more broadly would appear indicated.

7.4 Limitations

The second half of the thesis aimed to address specific technical questions regarding Toshiba 2D-SWE measurement variability and required measurement number. Patients in the cohort did not undergo liver biopsy, and the study was therefore not designed to evaluate the accuracy of Toshiba 2D-SWE. As a consequence, numerous questions remain for the new 2D-SWE technique and the tool still requires clinical validation in the quantification of liver fibrosis. Whilst our study does not evaluate these areas, assisting in

the development of acquisition protocols will hopefully facilitate the rigorous and standardized assessment of 2D-SWE performance in future clinical trials.

Another major limitation relates to the study's small cohort size (n=55), which limits the power of our analyses. The study's primary findings are nonetheless all drawn from results which are of high statistical significance. This further underscores the large effect sizes observed and therefore the likely clinical relevance of our study findings.

Finally, the small cohort size also raises the possibility of selection bias and uncertainties as to whether the Toshiba 2D-SWE cohort is representative of the true target population. The demographic, anthropometric measures and CLD aetiologies of the 2D-SWE cohort was however almost identical to the greater ARFI cohort. This is somewhat reassuring, as the ARFI cohort itself was felt to be very closely reflective of the overall target population (discussed previously).

Chapter 8. Conclusion

Elastography is increasingly relied upon as a clinical substitute for liver biopsy in the non-invasive assessment of fibrosis. Whilst ultrasound based SWE techniques continue to expand, current tools are imperfect and their clinical limitations remain incompletely understood. Ongoing efforts are therefore essential to help improve our understanding and effective utilisation of existing SWE elastography tools, but also to foster new technologies which may improve the accuracy of fibrosis quantification in the future. The above work contributes to the literature on both points.

ARFI (Siemens)

Our analyses for ARFI were wide-ranging, and aimed to shed light on a number of unresolved questions surrounding the point SWE technique.

Our results provide some of the first indications of local ARFI performance in a clinical Australian setting, with ARFI being almost exclusively validated in European and Asian cohorts to date. Whilst ARFI showed local utility in liver fibrosis quantification and to a lesser extent cirrhosis assessment, our results suggest it may not be performing at the high levels reported in the literature. The technique is particularly prone to false positive results, and clinicians need to take care in diagnosing advanced fibrosis on the basis of an elevated ARFI LSM. ARFI's main utility instead appears to be in the exclusion of liver fibrosis, with the technology exhibiting high sensitivity and NPV in our cohort.

We found obesity to be the primary factor impacting on ARFI performance, particularly when centrally distributed. The literature surrounding the relationship between body habitus and ARFI performance is inconsistent, and our results provide some of the strongest confirmation and warning regarding obesity's marked impact on all facets of ARFI performance. Our SLD analyses also provide insight into the mechanism underlying this relationship, supporting existing hypotheses implicating subcutaneous adipose tissue in the attenuation of the ARFI push pulse and / or tracking beams.

The other findings of potential high clinical utility includes the development and evaluation of two new reliability assessment strategies; scanning patients with multiple independent operators and SLD. Neither approach has been previously evaluated to our

knowledge, and both showed high utility in triaging the reliability of obtained ARFI results. Clinicians currently have very limited facility to assess ARFI reliability, which is one of the primary limitations of the point SWE technique. The two approaches may therefore allow operators to better assess the validity of obtained ARFI results, providing incremental information above the existing but imperfect IQR/Median criteria.

Our ARFI analyses are relatively unique in being drawn from one of the largest patient cohorts assessed to date, as well as for the large number of patients scanned with multiple operators. Conversely, our study was limited by the small number patients who completed a contemporaneous liver biopsy, which reduced the power and reliability of our accuracy analyses. As a consequence, some of our findings require further validation in a larger cohort of patients with a histopathologic reference. This includes further clarification of ARFI's low local accuracy, but also the validation of SLD and multiple operators as reliability metrics.

2D-SWE (Toshiba)

In contrast, the new Toshiba 2D-SWE technique is in its clinical infancy and remains minimally evaluated to date. Our Toshiba 2D-SWE analyses were accordingly narrow and aimed to address specific technical questions to assist in the formation of acquisition guidelines, and thereby provide a framework for future clinical studies and practice.

The primary study objective was to assess 2D-SWE measurement variability to help determine the number of measurements required per patient; and on this front acquiring five readings appeared sufficient for clinical purposes. Many of our ARFI findings were similarly observed with Toshiba 2D-SWE system, including the impact of obesity and subcapsular measurements on 2D-SWE reliability. This crossover is not unexpected given the shared technological principles underpinning the two systems, and suggests that many of the lessons learnt from point SWE may be transferrable to Toshiba 2D-SWE technique.

Finally, we also found ROI SD/Speed to have high potential value as a reliability indicator; allowing operators to assess the reliability of individual 2D-SWE measurements in real-time during the scan acquisition process. Whilst the relationship between ROI SD/Speed and Toshiba 2D-SWE accuracy requires assessment in future studies, the potential clinical

applications of ROI SD/Speed are wide-ranging. And ROI SD/Speed may furthermore hold relevance for related 2D-SWE devices, including SSI.

Concluding statement

In conclusion, the work encompassed in this thesis contributes to our knowledge of ultrasound SWE on two fronts. It firstly aims to improve our understanding and effective utilisation of existing SWE tools, but also looks forwards by helping to lay foundations for the emerging and more technologically advanced Toshiba 2D-SWE technique. The work will therefore hopefully contribute to the effective utilisation of these exciting technologies, both now and into the future.

References

1. Adams LA, Lymp JF, St. Sauver J, et al. The natural history of nonalcoholic fatty liver disease: A population-based cohort study. *Gastroenterology*. 2005;129(1):113-121. doi:10.1053/j.gastro.2005.04.014
2. Poynard T, Bedossa P, Opolon P. Natural history of liver fibrosis progression in patients with chronic hepatitis C. *Lancet*. 1997;349:825-832. doi:10.1016/S0140-6736(96)07642-8
3. Angulo P, Kleiner DE, Dam-Larsen S, et al. Liver fibrosis, but no other histologic features, is associated with long-term outcomes of patients with nonalcoholic fatty liver disease. *Gastroenterology*. 2015;149(2):389-397.e10. doi:10.1053/j.gastro.2015.04.043
4. Everhart JE, Wright EC, Goodman ZD, et al. Prognostic value of ishak fibrosis stage: Findings from the hepatitis C antiviral long-term treatment against cirrhosis trial. *Hepatology*. 2010;51(2):585-594. doi:10.1002/hep.23315
5. Bedossa P, Dargere D, Paradis V. Sampling variability of liver fibrosis in chronic hepatitis C. *Hepatology*. 2003;38:1449-1457. doi:10.1016/j.hep.2003.09.022
6. Maharaj B, Maharaj RJ, Leary WP, et al. Sampling variability and its influence on the diagnostic yield of percutaneous needle biopsy of the liver. *Lancet*. 1986;1:523-525. doi:10.1016/S0022-3468(87)80195-1
7. RG L. *Diagnostic Liver Pathology*. *St. Louis: Mosby-Year Book.*; 1994.
8. Janes CH, Lindor KD. Outcome of patients hospitalized for complications after outpatient liver biopsy. *Ann Intern Med*. 1993;118(2):96-98. doi:10.1059/0003-4819-118-2-199301150-00003
9. Rockey DC, Caldwell SH, Goodman ZD, Nelson RC, Smith AD. Liver biopsy. *Hepatology*. 2009;49(3):1017-1044. doi:10.1002/hep.22742
10. Nightingale K. Acoustic Radiation Force Impulse (ARFI) Imaging: a Review. *Curr Med Imaging Rev*. 2012;7(4):328-339. doi:10.2174/157340511798038657.Acoustic
11. Bota S, Sporea I, Sirli R, et al. Factors associated with the impossibility to obtain reliable liver stiffness measurements by means of Acoustic Radiation Force Impulse (ARFI) elastography - Analysis of a cohort of 1031 subjects. *Eur J Radiol*. 2014;83(2):268-272. doi:10.1016/j.ejrad.2013.11.019
12. Castera L, Foucher J, Bernard PH, et al. Pitfalls of liver stiffness measurement: A 5-year prospective study of 13,369 examinations. *Hepatology*. 2010;51:828-835. doi:10.1002/hep.23425
13. Sporea I, Sirli R, Bota S, Popescu A, Sendroiu M, Jurchis A. Comparative Study Concerning the Value of Acoustic Radiation Force Impulse Elastography (ARFI) in Comparison with Transient Elastography (TE) for the Assessment of Liver Fibrosis in Patients with Chronic Hepatitis B and C. *Ultrasound Med Biol*. 2012;38(8):1310-1316. doi:10.1016/j.ultrasmedbio.2012.03.011
14. Friedrich-Rust M, Wunder K, Kriener S, et al. Liver fibrosis in viral hepatitis: noninvasive assessment with acoustic radiation force impulse imaging versus transient elastography. *Radiology*. 2009;252(2):595-604. doi:10.1148/radiol.2523081928
15. Bota S, Herkner H, Sporea I, et al. Meta-analysis: ARFI elastography versus transient elastography for the evaluation of liver fibrosis. *Liver Int*. 2013;33(8):1138-1147. doi:10.1111/liv.12240
16. Chen SH, Li YF, Lai HC, et al. Effects of patient factors on noninvasive liver stiffness measurement using acoustic radiation force impulse elastography in patients with chronic hepatitis C. *BMC Gastroenterol*. 2012;12(2):105.

doi:http://dx.doi.org/10.1186/1471-230X-12-105

17. Bota S, Sporea I, Sirli R, et al. ARFI elastography vs. transient elastography: Which one is more influenced by high aminotransferases values. In: *Hepatology*. Vol 56. ; 2012:839A. doi:http://dx.doi.org/10.1002/hep.26040

18. Ebinuma H, Saito H, Komuta M, et al. Evaluation of liver fibrosis by transient elastography using acoustic radiation force impulse: comparison with Fibroscan(®). *J Gastroenterol*. 2011;46(10):1238-1248. doi:10.1007/s00535-011-0437-3

19. Goertz RS, Egger C, Neurath MF, Strobel D. Impact of Food Intake, Ultrasound Transducer, Breathing Maneuvers and Body Position on Acoustic Radiation Force Impulse (ARFI) Elastometry of the Liver. *Ultraschall der Medizin*. 2012;33(4):380-385. doi:10.1055/s-0032-1312816

20. Popescu A, Bota S, Sporea I, et al. The Influence of Food Intake on Liver Stiffness Values Assessed by Acoustic Radiation Force Impulse Elastography-Preliminary Results. *Ultrasound Med Biol*. 2013;39(4):579-584. doi:10.1016/j.ultrasmedbio.2012.11.013

21. Chang S, Kim M-J, Kim J, Lee M-J. Variability of shear wave velocity using different frequencies in acoustic radiation force impulse (ARFI) elastography: a phantom and normal liver study. *Ultraschall Med*. 2013;34(3):260-265. doi:10.1055/s-0032-1313008

22. Sporea I, Sirli RL, Deleanu A, et al. Acoustic radiation force impulse elastography as compared to transient elastography and liver biopsy in patients with chronic hepatopathies. *Ultraschall Med*. 2011;32 Suppl 1:S46-52. doi:10.1055/s-0029-1245360

23. Cassinotto C, Lapuyade B, Aït-Ali A, et al. Liver fibrosis: noninvasive assessment with acoustic radiation force impulse elastography--comparison with FibroScan M and XL probes and FibroTest in patients with chronic liver disease. *Radiology*. 2013;269:283-292. doi:10.1148/radiol.13122208

24. Cassinotto C, Boursier J, de L??dinghen V, et al. Liver stiffness in nonalcoholic fatty liver disease: A comparison of supersonic shear imaging, FibroScan, and ARFI with liver biopsy. *Hepatology*. 2016;63(6):1817-1827. doi:10.1002/hep.28394

25. Nishikawa T, Hashimoto S, Kawabe N, et al. Factors correlating with acoustic radiation force impulse elastography in chronic hepatitis C. *World J Gastroenterol*. 2014;20(5):1289-1297. doi:10.3748/wjg.v20.i5.1289

26. Park MS, Kim SW, Yoon KT, et al. Factors Influencing the Diagnostic Accuracy of Acoustic Radiation Force Impulse Elastography in Patients with Chronic Hepatitis B. *Gut Liver*. 2015;10(2):275-282. doi:10.5009/gnl14391

27. Palmeri ML, Wang MH, Rouze NC, et al. Noninvasive evaluation of hepatic fibrosis using acoustic radiation force-based shear stiffness in patients with nonalcoholic fatty liver disease. *J Hepatol*. 2011;55(3):666-672. doi:10.1016/j.jhep.2010.12.019

28. Attia D, Bantel H, Lenzen H, Manns MP, Gebel MJ, Potthoff A. Liver stiffness measurement using acoustic radiation force impulse elastography in overweight and obese patients. *Aliment Pharmacol Ther*. 2016;44(4):366-379. doi:10.1111/apt.13710

29. Guzmán-Aroca F, Frutos-Bernal MD, Bas A, et al. Detection of non-alcoholic steatohepatitis in patients with morbid obesity before bariatric surgery: preliminary evaluation with acoustic radiation force impulse imaging. *Eur Radiol*. 2012;22:2525-2532. doi:10.1007/s00330-012-2505-3

30. Gerstenmaier JF, Gibson RN. Ultrasound in chronic liver disease. *Insights Imaging*. 2014;5(4):441-455. doi:10.1007/s13244-014-0336-2

31. Bota S, Sporea I, Şirli R, Popescu A, Dănilă M, Şendroiu M. Factors that influence the correlation of acoustic radiation force impulse (ARFI) elastography with liver fibrosis. *Med Ultrason*. 2011;13(2):135-140.

32. Rifai K, Cornberg J, Mederacke I, et al. Clinical feasibility of liver elastography by

acoustic radiation force impulse imaging (ARFI). *Dig Liver Dis*. 2011;43(6):491-497. doi:10.1016/j.dld.2011.02.011

33. Ferraioli G, Tinelli C, Lissandrin R ZM. Ultrasound point shear wave elastography assessment of liver and spleen stiffness: effect of training on repeatability of measurements. *Eur Radiol*. 2014;24(6):1283-1289.

34. Boursier J, Isselin G, Fouchard-Hubert I, et al. *Acoustic Radiation Force Impulse: A New Ultrasonographic Technology for the Widespread Noninvasive Diagnosis of Liver Fibrosis*. Vol 22.; 2010. doi:10.1097/MEG.0b013e328339e0a1

35. 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 der Medizin*. 2017;38(4):e16-e47. doi:10.1055/s-0043-103952

36. Goertz RS, Sturm J, Pfeifer L, et al. ARFI cut-off values and significance of standard deviation for liver fibrosis staging in patients with chronic liver disease. *Ann Hepatol*. 2013;12(6):935-941. http://www.ncbi.nlm.nih.gov/pubmed/24114824.

37. Salzl P, Reiberger T, Ferlitsch M, et al. Evaluation of portal hypertension and varices by acoustic radiation force impulse imaging of the liver compared to transient elastography and AST to platelet ratio index. *Ultraschall Med*. 2014;35(6):528-533. doi:10.1055/s-0034-1366506

38. Attia D, Schoenemeier B, Rodt T, et al. Evaluation of liver and spleen stiffness with acoustic radiation force impulse quantification elastography for diagnosing clinically significant portal hypertension. *Ultraschall der Medizin*. 2015;36(6):603-610. doi:10.1055/s-0041-107971

39. Morishita N, Hiramatsu N, Oze T, et al. Liver stiffness measurement by acoustic radiation force impulse is useful in predicting the presence of esophageal varices or high-risk esophageal varices among patients with HCV-related cirrhosis. *J Gastroenterol*. 2014;49(7):1175-1182. doi:10.1007/s00535-013-0877-z

40. Park Y, Kim SU, Park SY, et al. A novel model to predict esophageal varices in patients with compensated cirrhosis using acoustic radiation force impulse elastography. *PLoS One*. 2015;10(3). doi:10.1371/journal.pone.0121009

41. Takuma Y, Nouso K, Morimoto Y, et al. Measurement of spleen stiffness by acoustic radiation force impulse imaging identifies cirrhotic patients with esophageal varices. *Gastroenterology*. 2013;144(1):92-101.e2. doi:10.1053/j.gastro.2012.09.049

42. Ye X, Ran H, Cheng J, et al. Liver and spleen stiffness measured by acoustic radiation force impulse elastography for noninvasive assessment of liver fibrosis and esophageal varices in patients with chronic hepatitis B. *J ultrasound Med Off J Am Inst Ultrasound Med*. 2012;31:1245-1253.

43. Sirli R, Sporea I, Bota SI, Ciof OP, Popescu A, Danila M. Can acoustic radiation force impulse elastography (ARFI) predict the complications of liver cirrhosis? *Gastroenterology*. 2010;138(5):359. http://www.embase.com/search/results?subaction=viewrecord&from=export&id=L70396570%5Cnhttp://sfx.library.uu.nl/utrecht?sid=EMBASE&issn=00165085&id=doi:&atitle=Can+acoustic+radiation+force+impulse+elastography+(ARFI)+predict+the+complications+of+liver+cirr.

44. Bota S, Sporea I, Sirli R, Popescu A, Dănilă M, Sendroiu M. The influence of liver residual mass on the values of Acoustic Radiation Force Impulse Elastography (ARFI) in cirrhotic patients. *Med Ultrason*. 2011;13:195-199.

45. Vermehren J, Polta A, Zimmermann O, et al. Comparison of acoustic radiation force impulse imaging with transient elastography for the detection of complications in patients with cirrhosis. *Liver Int*. 2012;32(5):852-858. doi:10.1111/j.1478-3231.2011.02736.x

46. 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®. *J Hepatol*. 2014;61(3):550-557. doi:10.1016/j.jhep.2014.04.044

47. Lee MS, Bae JM, Joo SK, Woo H, Lee DH, Jung YJ, Kim BG, Lee KL KW. Prospective comparison among transient elastography, supersonic shear imaging, and ARFI imaging for predicting fibrosis in nonalcoholic fatty liver disease. *PLoS One*. 2017;12(11):e0188321.

48. Friedman SL. Liver fibrosis – from bench to bedside. *J Hepatol*. 2003;38:38-53. doi:10.1016/S0168-8278(02)00429-4

49. Friedman SL. The cellular basis of hepatic fibrosis - mechanisms and treatment strategies. *N Engl J Med*. 1993;328(25):1828-1835. doi:10.1056/NEJM199306243282508

50. Wells RG, Schwabe R. Origin and function of myofibroblasts in the liver. *Semin Liver Dis*. 2015;35(2):97-106. doi:10.1055/s-0035-1550061

51. Jaeschke H. Inflammation in response to hepatocellular apoptosis. *Hepatology*. 2002;35(4):964-966. doi:10.1053/jhep.2002.0350964

52. Rojkind M, Giambrone MA, Biempica L. Collagen Types in Normal and Cirrhotic Liver. *Gastroenterology*. 1979;76(4):710-719. doi:10.1016/S0016-5085(79)80170-5

53. Kubota K. Measurement of liver volume and hepatic functional reserve as a guide to decision-making in resectional surgery for hepatic tumors. *Hepatology*. 1997;26(5):1176-1181. doi:10.1053/jhep.1997.v26.pm0009362359

54. Bataller R, Brenner D. Liver fibrosis. *J Clin Invest*. 2005;115(2):209-218. doi:10.1172/JCI200524282.The

55. Benyon RC, Iredale JP. Is liver fibrosis reversible? *Gut*. 2000;46(4):443-446. doi:10.1136/gut.46.4.443

56. George SL, Bacon BR, Brunt EM, Mihindukulasuriya KL, Hoffman J, Di Bisceglie AM. Clinical, virologic, histologic, and biochemical outcomes after successful HCV therapy: A 5-year follow-up of 150 patients. *Hepatology*. 2009;49(3):729-738. doi:10.1002/hep.22694

57. Chang TT, Liaw YF, Wu SS, et al. Long-term entecavir therapy results in the reversal of fibrosis/cirrhosis and continued histological improvement in patients with chronic hepatitis B. *Hepatology*. 2010;52(3):886-893. doi:10.1002/hep.23785

58. Dienstag JL, Goldin RD, Heathcote EJ, et al. Histological outcome during long-term lamivudine therapy. *Gastroenterology*. 2003;124(1):105-117. doi:10.1053/gast.2003.50013

59. Czaja AJ, Carpenter HA. Decreased fibrosis during corticosteroid therapy of autoimmune hepatitis. In: *Journal of Hepatology*. Vol 40. ; 2004:646-652. doi:10.1016/j.jhep.2004.01.009

60. Dixon JB, Bhathal PS, Hughes NR, O'Brien PE. Nonalcoholic fatty liver disease: Improvement in liver histological analysis with weight loss. *Hepatology*. 2004;39(6):1647-1654. doi:10.1002/hep.20251

61. Niederau C, Fischer R, Pürschel A, Stremmel W, Häussinger D, Strohmeyer G. Long-term survival in patients with hereditary hemochromatosis. *Gastroenterology*. 1996;110:1107-1119. doi:10.1053/gast.1996.v110.pm8613000

62. Dufour JF, Delellis R, Kaplan MM. Regression of hepatic fibrosis in hepatitis C with long-term interferon treatment. *Dig Dis Sci*. 1998;43(12):2573-2576. doi:10.1023/A:1026601904609

63. Wanless IR, Nakashima E, Sherman M. Regression of human cirrhosis: Morphologic

features and the genesis of incomplete septal cirrhosis. *Arch Pathol Lab Med*. 2000;124(11):1599-1607. doi:10.1043/0003-9985(2000)124<1599:ROHC>2.0.CO;2

64. Strader DB, Wright T, Thomas DL, Seeff LB. Diagnosis, Management, and Treatment of Hepatitis C. *Hepatology*. 2004;39(4):1147-1171. doi:10.1002/hep.20119

65. Castera L, Yuen Chan HL, Arrese M, et al. EASL-ALEH Clinical Practice Guidelines: Non-invasive tests for evaluation of liver disease severity and prognosis. *J Hepatol*. 2015;63(1):237-264. doi:10.1016/j.jhep.2015.04.006

66. D'Amico G, Garcia-Tsao G, Pagliaro L. Natural history and prognostic indicators of survival in cirrhosis: A systematic review of 118 studies. *J Hepatol*. 2006;44(1):217-231. doi:10.1016/j.jhep.2005.10.013

67. Hytiroglou P, Snover DC, Alves V, et al. Beyond "cirrhosis": a proposal from the International Liver Pathology Study Group. *Am J Clin Pathol*. 2012. doi:10.1309/AJCP2T2OHTAPBTMP

68. GBD 2013 Mortality and Causes of Death Collaborators. Global, regional, and national age-sex specific all-cause and cause-specific mortality for 240 causes of death, 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet (London, England)*. 2015;385(9963):117-171. doi:10.1016/S0140-6736(14)61682-2

69. Lazo M, Clark JM. The epidemiology of nonalcoholic fatty liver disease: A global perspective. *Semin Liver Dis*. 2008;28(4):339-350. doi:10.1055/s-0028-1091978

70. Vernon G, Baranova A, Younossi ZM. Systematic review: The epidemiology and natural history of non-alcoholic fatty liver disease and non-alcoholic steatohepatitis in adults. *Aliment Pharmacol Ther*. 2011;34(3):274-285. doi:10.1111/j.1365-2036.2011.04724.x

71. Colloredo G, Guido M, Sonzogni A, Leandro G. Impact of liver biopsy size on histological evaluation of chronic viral hepatitis: The smaller the sample, the milder the disease. *J Hepatol*. 2003;39(2):239-244. doi:10.1016/S0168-8278(03)00191-0

72. Demetris AJ, Ruppert K. Pathologist's perspective on liver needle biopsy size? *J Hepatol*. 2003;39(2):275-277. doi:10.1016/S0168-8278(03)00282-4

73. Bedossa et al. Intraobserver and interobserver variations in liver biopsy interpretation in patients with chronic hepatitis C. The French METAVIR Cooperative Study Group. *Hepatology*. 1994;20(1 Pt 1):15-20. doi:S0270913994002016 [pii]

74. Regev A, Berho M, Jeffers LJ, et al. Sampling error and intraobserver variation in liver biopsy in patients with chronic HCV infection. *Am J Gastroenterol*. 2002;97:2614-2618. doi:10.1111/j.1572-0241.2002.06038.x

75. Goodman ZD. Grading and staging systems for inflammation and fibrosis in chronic liver diseases. *J Hepatol*. 2007;47(4):598-607. doi:10.1016/j.jhep.2007.07.006

76. Piccinino F, Sagnelli E, Pasquale G, Giusti G. Complications following percutaneous liver biopsy. A multicentre retrospective study on 68,276 biopsies. *J Hepatol*. 1986;2(2):165-173. doi:http://dx.doi.org/10.1016/S0168-8278(86)80075-7

77. Castéra L, Nègre I, Samii K, Buffet C. Pain experienced during percutaneous liver biopsy. *Hepatology*. 1999;30(6):1529-1530. doi:10.1002/hep.510300624

78. Sebastiani G, Alberti A. Non invasive fibrosis biomarkers reduce but not substitute the need for liver biopsy. *World J Gastroenterol*. 2006;12(23):3682-3694. doi:10.3748/wjg.v12.i23.3682

79. Bosch J, Abraldes JG, Berzigotti A, García-Pagan JC. The clinical use of HVPG measurements in chronic liver disease. *Nat Rev Gastroenterol Hepatol*. 2009. doi:10.1038/nrgastro.2009.149

80. Martinez SM, Crespo G, Navasa M, Forns X. Noninvasive assessment of liver fibrosis. *Hepatology*. 2011;53(1):325-335. doi:10.1002/hep.24013

81. Poynard T, Imbert-Bismut F, Munteanu M, et al. Overview of the diagnostic value of biochemical markers of liver fibrosis (FibroTest, HCV FibroSure) and necrosis (ActiTest) in patients with chronic hepatitis C. *Comp Hepatol*. 2004;3(1):8. doi:10.1186/1476-5926-3-8

82. Adams LA, Bulsara M, Rossi E, et al. Hepascore: an accurate validated predictor of liver fibrosis in chronic hepatitis C infection. *Clin Chem*. 2005;51(10):1867-1873. doi:10.1373/clinchem.2005.048389

83. Patel K, Gordon SC, Jacobson I, et al. Evaluation of a panel of non-invasive serum markers to differentiate mild from moderate-to-advanced liver fibrosis in chronic hepatitis C patients. *J Hepatol*. 2004;41(6):935-942. doi:10.1016/j.jhep.2004.08.008

84. Lichtinghagen R, Pietsch D, Bantel H, Manns MP, Brand K, Bahr MJ. The Enhanced Liver Fibrosis (ELF) score: Normal values, influence factors and proposed cut-off values. *J Hepatol*. 2013;59(2):236-242. doi:10.1016/j.jhep.2013.03.016

85. Wai CT, Greenson JK, Fontana RJ, et al. A simple noninvasive index can predict both significant fibrosis and cirrhosis in patients with chronic hepatitis C. *Hepatology*. 2003;38(2):518-526. doi:10.1053/jhep.2003.50346

86. Sterling RK, Lissen E, Clumeck N, et al. Development of a simple noninvasive index to predict significant fibrosis in patients with HIV/HCV coinfection. *Hepatology*. 2006;43(6):1317-1325. doi:10.1002/hep.21178

87. Sheth SG, Flamm SL, Gordon FD, Chopra S. AST/ALT ratio predicts cirrhosis in patients with chronic hepatitis C virus infection. *Am J Gastroenterol*. 1998;93(1):44-48. doi:10.1111/j.1572-0241.1998.044_c.x

88. Rossi E, Adams L a, Bulsara M, Jeffrey GP. Assessing liver fibrosis with serum marker models. *Clin Biochem Rev*. 2007;28(1):3-10.

89. Parkes J, Guha IN, Roderick P, Rosenberg W. Performance of serum marker panels for liver fibrosis in chronic hepatitis C. *J Hepatol*. 2006;44(3):462-474. doi:10.1016/j.jhep.2005.10.019

90. Bissell DM. Assessing fibrosis without a liver biopsy: Are we there yet? *Gastroenterology*. 2004;127(6):1847-1849. doi:10.1053/j.gastro.2004.10.012

91. Garra BS. Imaging and estimation of tissue elasticity by ultrasound. *Ultrasound Q*. 2007;23(4):255-268. doi:10.1097/ruq.0b013e31815b7ed6

92. Greenleaf JF, Fatemi M, Insana M. Selected Methods for Imaging Elastic Properties of Biological Tissues. *Annu Rev Biomed Eng*. 2003;5(1):57-78. doi:10.1146/annurev.bioeng.5.040202.121623

93. Ferraioli G, Filice C, Castera L, et al. WFUMB Guidelines and Recommendations for Clinical Use of Ultrasound Elastography: Part 3: Liver. *Ultrasound Med Biol*. 2015;41(5):1161-1179. doi:10.1016/j.ultrasmedbio.2015.03.007

94. Horster S, Mandel P, Zachoval R, Clevert DA. Comparing acoustic radiation force impulse imaging to transient elastography to assess liver stiffness in healthy volunteers with and without valsalva manoeuvre. *Clin Hemorheol Microcirc*. 2010;46:159-168. doi:10.3233/CH-2010-1342

95. 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 Gastroenterol*. 2011;46(12):1458-1467. doi:10.3109/00365521.2011.610004

96. Millonig G, Reimann FM, Friedrich S, et al. Extrahepatic cholestasis increases liver stiffness (FibroScan) irrespective of fibrosis. *Hepatology*. 2008;48:1718-1723. doi:10.1055/s-0029-1191982

97. Kwak JY, Kim E-K. Ultrasound elastography for thyroid nodules: recent advances.

Ultrason (Seoul, Korea). 2014;33(2):75-82. doi:10.14366/usg.13025

98. Palmeri ML, Sharma AC, Bouchard RR, Nightingale RW, Nightingale KR. A finite-element method model of soft tissue response to impulsive acoustic radiation force. *IEEE Trans Ultrason Ferroelectr Freq Control*. 2005;52(10):1699-1712. doi:10.1109/TUFFC.2005.1561624

99. Nightingale KR, Soo MS, Nightingale RW, Trahey GE. Acoustic radiation force impulse imaging: In vivo demonstration of clinical feasibility. *Ultrasound Med Biol*. 2002;28(2):227-235. doi:10.1016/S0301-5629(01)00499-9

100. Gennisson JL, Deffieux T, Fink M, Tanter M. Ultrasound elastography: Principles and techniques. *Diagn Interv Imaging*. 2013;94(5):487-495. doi:10.1016/j.diii.2013.01.022

101. Sarvazyan AP, Rudenko O V, Swanson SD, Fowlkes JB, Emelianov SY. Shear Wave Elasticity Imaging : A new ultrasonic technology of medical diagnostic. *Ultrasound Med Biol*. 1998;24(9):1419-1435. doi:10.1016/S0301-5629(98)00110-0

102. Dillman JR, Chen S, Davenport MS, et al. Superficial ultrasound shear wave speed measurements in soft and hard elasticity phantoms: repeatability and reproducibility using two ultrasound systems. *Pediatr Radiol*. 2015;45(3):376-385. doi:10.1007/s00247-014-3150-6

103. Bruno C, Minniti S, Bucci A, Pozzi Mucelli R. ARFI: from basic principles to clinical applications in diffuse chronic disease—a review. *Insights Imaging*. 2016;7(5):735-746. doi:10.1007/s13244-016-0514-5

104. Nightingale K. Acoustic Radiation Force Impulse (ARFI) Imaging: A Review. *Curr Med Imaging Rev*. 2011;7:328-339. doi:10.2174/157340511798038657

105. Dickinson RJ, Hill CR. Measurement of soft tissue motion using correlation between A-scans. *Ultrasound Med Biol*. 1982;8(3):263-271. doi:10.1016/0301-5629(82)90032-1

106. Castera L, Forns X, Alberti A. Non-invasive evaluation of liver fibrosis using transient elastography. *J Hepatol*. 2008;48(5):835-847. doi:10.1016/j.jhep.2008.02.008

107. Sandrin L, Fourquet B, Hasquenoph J-M, et al. Transient elastography: a new noninvasive method for assessment of hepatic fibrosis. *Ultrasound Med Biol*. 2003;29(12):1705-1713. doi:10.1016/j.ultrasmedbio.2003.07.001

108. Frulio N, Trillaud H. Ultrasound elastography in liver. *Diagn Interv Imaging*. 2013;94(5):515-534.

109. Ziol M, Handra-Luca A, Kettaneh A, et al. Noninvasive assessment of liver fibrosis by measurement of stiffness in patients with chronic hepatitis C. *Hepatology*. 2005;41(1):48-54. doi:10.1002/hep.20506

110. Stebbing J, Farouk L, Panos G, et al. A meta-analysis of transient elastography for the detection of hepatic fibrosis. *J Clin Gastroenterol*. 2010;44(3):214-219. doi:10.1097/MCG.0b013e3181b4af1f

111. Chon YE, Choi EH, Song KJ, et al. Performance of Transient Elastography for the Staging of Liver Fibrosis in Patients with Chronic Hepatitis B: A Meta-Analysis. *PLoS One*. 2012;7(9). doi:10.1371/journal.pone.0044930

112. Xu X, Su Y, Song R, Sheng Y, Ai W, Wu X. Performance of transient elastography assessing fibrosis of single hepatitis B virus infection: a systematic review and meta-analysis of a diagnostic test. *Hepatol Int*. 2015;9(4):558-566. doi:10.1007/s12072-015-9643-z

113. Kwok R, Tse Y-K, Wong GL-H, et al. Systematic review with meta-analysis: non-invasive assessment of non-alcoholic fatty liver disease--the role of transient elastography and plasma cytokeratin-18 fragments. *Aliment Pharmacol Ther*. 2014;39(3):254-269. doi:10.1111/apt.12569

114. Friedrich-Rust M, Ong MF, Martens S, et al. Performance of Transient Elastography for the Staging of Liver Fibrosis: A Meta-Analysis. *Gastroenterology*. 2008;134(4). doi:10.1053/j.gastro.2008.01.034

115. Robic MA, Procopet B, Metivier S, et al. Liver stiffness accurately predicts portal hypertension related complications in patients with chronic liver disease: a prospective study. *J Hepatol*. 2011;55(5):1017-1024. doi:10.1016/j.jhep.2011.01.051

116. Masuzaki R, Tateishi R, Yoshida H, et al. Prospective risk assessment for hepatocellular carcinoma development in patients with chronic hepatitis C by transient elastography. *Hepatology*. 2009;49(6):1954-1961. doi:10.1002/hep.22870

117. Jung KS, Kim SU, Ahn SH, et al. Risk assessment of hepatitis B virus-related hepatocellular carcinoma development using liver stiffness measurement (FibroScan). *Hepatology*. 2011;53(3):885-894. doi:10.1002/hep.24121

118. de Ledinghen V, Vergniol J, Barthe C, et al. Non-invasive tests for fibrosis and liver stiffness predict 5-year survival of patients chronically infected with hepatitis B virus. *Aliment Pharmacol Ther*. 2013;37(10):979-988. doi:10.1111/apt.12307

119. Vergniol J, Foucher J, Terrebonne E, et al. Noninvasive tests for fibrosis and liver stiffness predict 5-year outcomes of patients with chronic hepatitis C. *Gastroenterology*. 2011;140(7):1970-1979. doi:10.1053/j.gastro.2011.02.058

120. Yada N, Sakurai T, Minami T, et al. Ultrasound elastography correlates treatment response by antiviral therapy in patients with chronic hepatitis C. *Oncology*. 2014;87(suppl 1):118-123. doi:http://dx.doi.org/10.1159/000368155

121. Mandorfer M, Kozbial K, Freissmuth C, et al. Interferon-free regimens for chronic hepatitis C overcome the effects of portal hypertension on virological responses. *Aliment Pharmacol Ther*. 2015;42(6):707-718. doi:10.1111/apt.13315

122. Hezode C, Castera L, Roudot-Thoraval F, et al. Liver stiffness diminishes with antiviral response in chronic hepatitis C. *Aliment Pharmacol Ther*. 2011;34(6):656-663. doi:10.1111/j.1365-2036.2011.04765.x

123. Kim MN, Kim SU, Kim BK, et al. Long-term changes of liver stiffness values assessed using transient elastography in patients with chronic hepatitis B receiving entecavir. *Liver Int*. 2014;34(8):1216-1223. doi:http://dx.doi.org/10.1111/liv.12377

124. Enomoto M, Mori M, Ogawa T, et al. Usefulness of transient elastography for assessment of liver fibrosis in chronic hepatitis B: Regression of liver stiffness during entecavir therapy. *Hepatol Res*. 2010;40(9):853-861. doi:10.1111/j.1872-034X.2010.00687.x

125. Fung J, Lai C-L, Wong DK-H, Seto W-K, Hung I, Yuen M-F. Significant changes in liver stiffness measurements in patients with chronic hepatitis B: 3-year follow-up study. *J Viral Hepat*. 2011;18(7):e200-5. doi:10.1111/j.1365-2893.2010.01428.x

126. Lucidarme D, Foucher J, Le Bail B, et al. Factors of accuracy of transient elastography (fibroscan) for the diagnosis of liver fibrosis in chronic hepatitis C. *Hepatology*. 2009;49(4):1083-1089. doi:http://dx.doi.org/10.1002/hep.22748

127. Myers RP, Crotty P, Pomier-Layrargues G, Ma M, Urbanski SJ, Elkashab M. Prevalence, risk factors and causes of discordance in fibrosis staging by transient elastographyand liver biopsy. *Liver Int*. 2010;30(10):1471-1480. doi:10.1111/j.1478-3231.2010.02331.x

128. Boursier J, de Ledinghen V, Zarski JP, et al. Comparison of eight diagnostic algorithms for liver fibrosis in hepatitis C: New algorithms are more precise and entirely noninvasive. *Hepatology*. 2012;55(1):58-67. doi:10.1002/hep.24654

129. Boursier J, Zarski JP, de Ledinghen V, et al. Determination of reliability criteria for liver stiffness evaluation by transient elastography. *Hepatology*. 2013;57(3):1182-1191. doi:10.1002/hep.25993

130. Fraquelli M, Rigamonti C, Casazza G, et al. Reproducibility of transient elastography in the evaluation of liver fibrosis in patients with chronic liver disease. *Gut*. 2007;56(7):968-973. doi:10.1136/gut.2006.111302

131. Kemp W, Roberts S. Feasibility and performance of the FibroScan XL probe. *Hepatology*. 2012;55(4):1308-1309. doi:10.1002/hep.25521

132. Millonig G, Friedrich S, Adolf S, et al. Liver stiffness is directly influenced by central venous pressure. *J Hepatol*. 2010;52(2):206-210. doi:10.1016/j.jhep.2009.11.018

133. Chan HLY, Wong GLH, Choi PCL, et al. Alanine aminotransferase-based algorithms of liver stiffness measurement by transient elastography (Fibroscan) for liver fibrosis in chronic hepatitis B. *J Viral Hepat*. 2009;16(1):36-44. doi:10.1111/j.1365-2893.2008.01037.x

134. Coco B, Oliveri F, Maina AM, et al. Transient elastography: A new surrogate marker of liver fibrosis influenced by major changes of transaminases. *J Viral Hepat*. 2007;14(5):360-369. doi:10.1111/j.1365-2893.2006.00811.x

135. Muthupillai R, Lomas D, Rossman P, Greenleaf J, Manduca A, Ehman R. Magnetic resonance elastography by direct visualization of propagating acoustic strain waves. *Science (80-)*. 1995;269(5232):1854-1857. doi:10.1126/science.7569924

136. Wang Q-B, Zhu H, Liu H-L, Zhang B. Performance of magnetic resonance elastography and diffusion-weighted imaging for the staging of hepatic fibrosis: A meta-analysis. *Hepatology*. 2012;56(1):239-247. doi:10.1002/hep.25610

137. Singh S, Venkatesh SK, Wang Z, et al. Diagnostic performance of magnetic resonance elastography in staging liver fibrosis: A systematic review and meta-analysis of individual participant data. *Clin Gastroenterol Hepatol*. 2015;13(3):440-451. doi:10.1016/j.cgh.2014.09.046

138. Huwart L, Sempoux C, Vicaut E, et al. Magnetic Resonance Elastography for the Noninvasive Staging of Liver Fibrosis. *Gastroenterology*. 2008;135(1):32-40. doi:10.1053/j.gastro.2008.03.076

139. Chen J, Talwalkar JA, Yin M, et al. Assessment of clinical effectiveness of MR elastography and vibration-controlled transient elastography for detecting hepatic fibrosis. *Hepatology*. 2014;60:204A-205A. doi:http://dx.doi.org/10.1002/hep.27450

140. Barr RG, Ferraioli G, Palmeri ML, et al. Elastography Assessment of Liver Fibrosis: Society of Radiologists in Ultrasound Consensus Conference Statement. *Radiology*. 2015;32(0):94-107. doi:10.1097/RUQ.0000000000000209

141. Nightingale KR, Palmeri ML, Nightingale RW, Trahey GE. On the feasibility of remote palpation using acoustic radiation force. *J Acoust Soc Am*. 2001;110:625-634. doi:10.1121/1.1378344

142. Fahey BJ, Nightingale KR, Nelson RC, Palmeri ML, Trahey GE. Acoustic radiation force impulse imaging of the abdomen: Demonstration of feasibility and utility. *Ultrasound Med Biol*. 2005;31:1185-1198. doi:10.1016/j.ultrasmedbio.2005.05.004

143. Palmeri ML, Wang MH, Dahl JJ, Frinkley KD, Nightingale KR. Quantifying hepatic shear modulus in vivo using acoustic radiation force. *Ultrasound Med Biol*. 2008;34:546-558. doi:S0301-5629(07)00526-1 [pii]\r10.1016/j.ultrasmedbio.2007.10.009

144. D'Onofrio M, Crosara S, De Robertis R, et al. Acoustic radiation force impulse of the liver. *World J Gastroenterol*. 2013;19(30):4841-4849. doi:10.3748/wjg.v19.i30.4841

145. Nierhoff J, Lupsor M, Sporea I, et al. Performance of Acoustic Radiation Force Impulse imaging for the staging of liver fibrosis: a pooled meta-analysis. *J Viral Hepat*. 2012;19:e212-9. doi:10.1111/j.1365-2893.2011.01537.x

146. Sporea I, Bota S, Peck-Radosavljevic M, et al. Acoustic Radiation Force Impulse

Elastography for fibrosis evaluation in patients with chronic hepatitis C: An international multicenter study. *Eur J Radiol*. 2012;81(12):4112-4118. doi:10.1016/j.ejrad.2012.08.018

147. Rizzo L, Calvaruso V, Cacopardo B, et al. Comparison of Transient Elastography and Acoustic Radiation Force Impulse for Non-Invasive Staging of Liver Fibrosis in Patients With Chronic Hepatitis C. *Am J Gastroenterol*. 2011;106(12):2112-2120. doi:10.1038/ajg.2011.341

148. Yoneda M, Suzuki K, Kato S, et al. Nonalcoholic fatty liver disease: US-based acoustic radiation force impulse elastography. *Radiology*. 2010;256(2):640-647. doi:10.1148/radiol.10091662

149. Dong DR, Hao MN, Li C, et al. Acoustic radiation force impulse elastography, FibroScan??, Forns' index and their combination in the assessment of liver fibrosis in patients with chronic hepatitis B, and the impact of inflammatory activity and steatosis on these diagnostic methods. *Mol Med Rep*. 2015;11(6):4174-4182. doi:10.3892/mmr.2015.3299

150. Friedrich-Rust M, Romen D, Vermehren J, et al. Acoustic radiation force impulse-imaging and transient elastography for non-invasive assessment of liver fibrosis and steatosis in NAFLD. *Eur J Radiol*. 2012;81(3):e325-31. doi:10.1016/j.ejrad.2011.10.029

151. Fierbinteanu Braticevici C, Sporea I, Panaitescu E, Tribus L. Value of acoustic radiation force impulse imaging elastography for non-invasive evaluation of patients with nonalcoholic fatty liver disease. *Ultrasound Med Biol*. 2013;39(11):1942-1950. doi:10.1016/j.ultrasmedbio.2013.04.019

152. Liu H, Fu J, Hong R, Liu L, Li F. Acoustic radiation force impulse elastography for the non-invasive evaluation of hepatic fibrosis in non-alcoholic fatty liver disease patients: A systematic review & meta-analysis. *PLoS One*. 2015;10(7). doi:10.1371/journal.pone.0127782

153. Kiani A, Brun V, Lainé F, et al. Acoustic radiation force impulse imaging for assessing liver fibrosis in alcoholic liver disease. *World J Gastroenterol*. 2016;22(20):4926-4935. doi:10.3748/wjg.v22.i20.4926

154. Zhang D, Li P, Chen M, et al. Non-invasive assessment of liver fibrosis in patients with alcoholic liver disease using acoustic radiation force impulse elastography. *Abdom Imaging*. 2014;40(4):723-729. doi:10.1007/s00261-014-0154-5

155. Zhang D-K, Chen M, Liu Y, Wang R-F, Liu L-P, Li M. Acoustic radiation force impulse elastography for non-invasive assessment of disease stage in patients with primary biliary cirrhosis: A preliminary study. *Clin Radiol*. 2014;69(8):836-840. doi:10.1016/j.crad.2014.03.019

156. Efe C, Gungoren MS, Ozaslan E, Akbiyik F, Kav T. Acoustic Radiation Force Impulse (ARFI) for Fibrosis Staging in Patients with Autoimmune Hepatitis. *Hepatogastroenterology*. 2015;62(139):670-672. http://www.ncbi.nlm.nih.gov/pubmed/26897951.

157. Hanquinet S, Courvoisier DS, Rougemont AL, et al. Contribution of acoustic radiation force impulse (ARFI) elastography to the ultrasound diagnosis of biliary atresia. *Pediatr Radiol*. 2015;45(10):1489-1495. doi:10.1007/s00247-015-3352-6

158. Bota S, Sporea I, Sirli R, Popescu A, Danila M, Costachescu D. Intra- and Interoperator Reproducibility of Acoustic Radiation Force Impulse (ARFI) Elastography-Preliminary Results. *Ultrasound Med Biol*. 2012;38(7):1103-1108. doi:10.1016/j.ultrasmedbio.2012.02.032

159. Ferraioli G, Tinelli C, Lissandrin R, et al. Point shear wave elastography method for assessing liver stiffness. *World J Gastroenterol*. 2014;20(16):4787-4796. doi:10.3748/wjg.v20.i16.4787

160. D'Onofrio M, Gallotti A, Mucelli RP, Pozzi Mucelli R. Tissue quantification with acoustic radiation force impulse imaging: Measurement repeatability and normal values in the healthy liver. *Am J Roentgenol*. 2010;195(1):132-136. doi:10.2214/AJR.09.3923

161. Guzmán-Aroca F, Reus M, Berná-Serna JD, et al. Reproducibility of shear wave velocity measurements by acoustic radiation force impulse imaging of the liver: a study in healthy volunteers. *J Ultrasound Med*. 2011;30(7):975-979.

162. Goertz RS, Amann K, Heide R, Bernatik T, Neurath MF, Strobel D. An abdominal and thyroid status with Acoustic Radiation Force Impulse Elastometry - A feasibility study: Acoustic Radiation Force Impulse Elastometry of human organs. *Eur J Radiol*. 2011;80(3). doi:10.1016/j.ejrad.2010.09.025

163. Bota S, Sporea I, Sirli R, et al. Can ARFI elastography predict the presence of significant esophageal varices in newly diagnosed cirrhotic patients? *Ann Hepatol*. 2012;11:519-525. http://www.ncbi.nlm.nih.gov/pubmed/22700634.

164. Ma X, Wang L, Wu H, et al. Spleen stiffness is superior to liver stiffness for predicting esophageal varices in chronic liver disease: A meta-analysis. *PLoS One*. 2016;11(11). doi:10.1371/journal.pone.0165786

165. Şirli R, Sporea I, Popescu A, Danila M. Ultrasound-based elastography for the diagnosis of portal hypertension in cirrhotics. *World J Gastroenterol*. 2015. doi:10.3748/wjg.v21.i41.11542

166. Kaminuma C, Tsushima Y, Matsumoto N, Kurabayashi T, Taketomi-Takahashi A, Endo K. Reliable measurement procedure of virtual touch tissue quantification with acoustic radiation force impulse imaging. *J Ultrasound Med*. 2011;30(6):745-751.

167. Yoon KT, Lim SM, Park JY, et al. Liver stiffness measurement using acoustic radiation force impulse (ARFI) elastography and effect of necroinflammation. *Dig Dis Sci*. 2012;57(6):1682-1691. doi:10.1007/s10620-012-2044-4

168. Bota S. Value of acoustic radiation force impulse elastography for the assessment of ascites syndrome. *World J Radiol*. 2011. doi:10.4329/wjr.v3.i8.205

169. Karlas T, Dietrich A, Peter V, et al. Evaluation of transient elastography, Acoustic Radiation Force Impulse Imaging (ARFI), and Enhanced Liver Function (ELF) score for detection of fibrosis in morbidly obese patients. *PLoS One*. 2015;10(11). doi:10.1371/journal.pone.0141649

170. Palmeri ML, Nightingale KR. Acoustic radiation force-based elasticity imaging methods. *Interface Focus*. 2011;1(4):553-564. doi:10.1098/rsfs.2011.0023

171. Bota S, Sporea I, Sirli R, Popescu A, Dănilă M, Sendroiu M. Factors that influence the correlation of acoustic radiation force impulse (ARFI), elastography with liver fibrosis. *Med Ultrason*. 2011;13(2):135-140. http://www.ncbi.nlm.nih.gov/pubmed/21655540.

172. Lupsor M, Badea R, Stefanescu H, et al. 404 PERFORMANCE OF ACOUSTIC RADIATION FORCE IMPULSE IMAGING IN THE NONINVASIVE ASSESSMENT OF LIVER FIBROSIS, IN COMPARISON TO UNIDIMENSIONAL TRANSIENT ELASTOGRAPHY. *J Hepatol*. 2010;52:S165-S166. doi:10.1016/S0168-8278(10)60406-0

173. Guo Y, Lin H, Zhang X, Wen H, Chen S CX. The influence of hepatic steatosis on the evaluation of fibrosis with non-alcoholic fatty liver disease by acoustic radiation force impulse. *Conf Proc IEEE Eng Med Biol Soc*. 2017:298802991.

174. Guzmán Aroca F, Ayala I, Serrano L, et al. Assessment of liver steatosis in chicken by using acoustic radiation force impulse imaging: Preliminary results. *Eur Radiol*. 2010;20(10):2367-2371. doi:10.1007/s00330-010-1808-5

175. Bota S, Sporea I, Sirli R, Popescu A, Jurchis A. Factors Which Influence the Accuracy of Acoustic Radiation Force Impulse (ARFI) Elastography for the Diagnosis of Liver Fibrosis

in Patients with Chronic Hepatitis C. *Ultrasound Med Biol*. 2013;39:407-412. doi:10.1016/j.ultrasmedbio.2012.09.017

176. Bavu É, Gennisson JL, Couade M, et al. Noninvasive In Vivo Liver Fibrosis Evaluation Using Supersonic Shear Imaging: A Clinical Study on 113 Hepatitis C Virus Patients. *Ultrasound Med Biol*. 2011;37(9):1361-1373. doi:10.1016/j.ultrasmedbio.2011.05.016

177. Muller M, Gennisson JL, Deffieux T, Tanter M, Fink M. Quantitative Viscoelasticity Mapping of Human Liver Using Supersonic Shear Imaging: Preliminary In Vivo Feasability Study. *Ultrasound Med Biol*. 2009;35(2):219-229. doi:10.1016/j.ultrasmedbio.2008.08.018

178. Ferraioli G, Tinelli C, Dal Bello B, Zicchetti M, Filice G, Filice C. Accuracy of real-time shear wave elastography for assessing liver fibrosis in chronic hepatitis C: a pilot study. *Hepatology*. 2012;56(6):2125-2133. doi:10.1002/hep.25936

179. Hall TJ, Milkowski A, Garra B, et al. RSNA/QIBA: Shear wave speed as a biomarker for liver fibrosis staging. In: *IEEE International Ultrasonics Symposium, IUS*. ; 2013:397-400. doi:10.1109/ULTSYM.2013.0103

180. Bercoff J, Tanter M, Fink M. Supersonic Shear Imaging : A New Technique. *IEEE Trans Ultrason Ferroelectr Freq Control*. 2004;51(4):396-409. doi:10.1109/TUFFC.2004.1295425

181. Gennisson J-L, Rénier M, Catheline S, et al. Acoustoelasticity in soft solids: Assessment of the nonlinear shear modulus with the acoustic radiation force. *J Acoust Soc Am*. 2007;122(6):3211-3219. doi:10.1121/1.2793605

182. Piscaglia F, Marinelli S, Bota S, et al. The role of ultrasound elastographic techniques in chronic liver disease: Current status and future perspectives. *Eur J Radiol*. 2013. doi:10.1016/j.ejrad.2013.06.009

183. Staugaard B, Christensen PB, Mössner B, et al. Feasibility of transient elastography versus real-time two-dimensional shear wave elastography in difficult-to-scan patients. *Scand J Gastroenterol*. 2016;51(11):1354-1359. doi:10.1080/00365521.2016.1193217

184. Leung VY, Shen J, Wong VW, et al. Quantitative Elastography of Liver Fibrosis and Spleen Stiffness in Chronic Hepatitis B Carriers: Comparison of Shear-Wave Elastography and Transient Elastography with Liver Biopsy Correlation. *Radiology*. 2013;269(3):910-918. doi:10.1148/radiol.13130128

185. Hudson JM, Milot L, Parry C, Williams R, Burns PN. Inter- and Intra-Operator Reliability and Repeatability of Shear Wave Elastography in the Liver: A Study in Healthy Volunteers. *Ultrasound Med Biol*. 2013;39(6):950-955. doi:10.1016/j.ultrasmedbio.2012.12.011

186. Sporea I, Grădinaru-Taşcău 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(4):268-272. doi:10.11152/mu.2013.2066.154.isp2

187. Wang CZ, Zheng J, Huang ZP, et al. Influence of Measurement Depth on the Stiffness Assessment ofHealthy Liver with Real-Time Shear Wave Elastography. *Ultrasound Med Biol*. 2014;40(3):461-469. doi:10.1016/j.ultrasmedbio.2013.10.021

188. Gradinaru-Tascau O, Sporea I, Bota S, et al. Does experience play a role in the ability to perform liver stiffness measurements by means of supersonic shear imaging (SSI)? *Med Ultrason*. 2013;15(3):180-183. doi:10.11152/mu.2013.2066.153.ogt1is2

189. Iijima H. Approaches to the diagnosis of Liver Fibrosis. *Toshiba Med Syst Corp - White Pap*. 2015.

190. WHO. Physical status: the use and interpretation of anthropometry. Report of a WHO Expert Committee. *World Health Organ Tech Rep Ser*. 1995;854:1-452. doi:854

191. Bellentani S, Tiribelli C, Saccoccio G, et al. Prevalence of chronic liver disease in the general population of northern Italy: the Dionysos Study. *Hepatology*. 1994;20(6):1442-1449. doi:10.1002/hep.1840200611

192. Kamath PS, Wiesner RH, Malinchoc M, et al. A model to predict survival in patients with end-stage liver disease. *Hepatology*. 2001;33(2):464-470. doi:10.1053/jhep.2001.22172

193. Wiesner R, Edwards E, Freeman R, et al. Model for end-stage liver disease (MELD) and allocation of donor livers. *Gastroenterology*. 2003;124(1):91-96. doi:10.1053/gast.2003.50016

194. Child CG, Turcotte JG, Child CG TJ. *The Liver and Portal Hypertension*.; 1964.

195. Vilstrup H, Amodio P, Bajaj J, Cordoba J, Ferenci P, Mullen KD, Weisseborn K WP. Hepatic encphalopathy in chronic liver disease: 2014 Practice Gudeline by the American Asociation for the Study of Liver Diseases and the European Association for the STudy of Liver. *Hepatology*. 2014;60(2):715.

196. Bruix J, Sherman M. Management of hepatocellular carcinoma: An update. *Hepatology*. 2011;53(3):1020-1022. doi:10.1002/hep.24199

197. Dufour JF, Greten TF, Raymond E, et al. Clinical Practice Guidelines EASL – EORTC Clinical Practice Guidelines : Management of hepatocellular carcinoma European Organisation for Research and Treatment of Cancer. *J Hepatol*. 2012;56(4):908-943. doi:10.1016/j.jhep.2011.12.001

198. Bedossa P, Poynard T. An algorithm for the grading of activity in chronic hepatitis C. The METAVIR Cooperative Study Group. *Hepatology*. 1996;24:289-293. doi:S0270913996003205 [pii]\r10.1002/hep.510240201

199. Scheuer PJ. Classification of chronic viral hepatitis: a need for reassessment. *J Hepatol*. 1991. doi:10.1016/0168-8278(91)90084-O

200. Ishak K, Baptista A, Bianchi L, et al. Histological grading and staging of chronic hepatitis. *J Hepatol*. 1995. doi:10.1016/0168-8278(95)80226-6

201. Brunt EM, Janney CG, Di Bisceglie AM, Neuschwander-Tetri BA, Bacon BR. Nonalcoholic steatohepatitis: A proposal for grading and staging the histological lesions. *Am J Gastroenterol*. 1999;94(9):2467-2474. doi:10.1111/j.1572-0241.1999.01377.x

202. Cronbach LJ. Coefficient alpha and the internal structure of tests. *Psychometrika*. 1951;16(3):297-334. doi:10.1007/BF02310555

203. Bland JM, Altman DG. Statistical Methods for Assessing Agreement Between Two Methods of Clinical Measurement. *Lancet*. 1986;327:307-310. doi:10.1016/S0140-6736(86)90837-8

204. Yoon JH, Lee JM, Han JK, Cho BI. Shear Wave Elastography for Liver Stiffness Measurement in Clinical Sonographic Examinations. *J Ultrasound Med*. 2014;33(3):437-447. doi:10.7863/ultra.33.3.437

205. Nierhoff J, Chávez Ortiz AA, Herrmann E, Zeuzem S, Friedrich-Rust M. The efficiency of acoustic radiation force impulse imaging for the staging of liver fibrosis: A meta-analysis. *Eur Radiol*. 2013;23(11):3040-3053. doi:10.1007/s00330-013-2927-6

206. Matteoni C, Younossi Z, Gramlich T, Boparai N, Liu Y, McCullough A. Nonalcoholic fatty liver disease: A spectrum of clinical and pathological severity☆, □□. *Gastroenterology*. 1999;116(6):1413-1419. doi:10.1016/S0016-5085(99)70506-8

207. Roulot D, Costes J-L, Buyck J-F, et al. Transient elastography as a screening tool for liver fibrosis and cirrhosis in a community-based population aged over 45 years. *Gut*. 2011;60(7):977-984. doi:10.1136/gut.2010.221382

208. Koehler EM, Plompen EPC, Schouten JNL, et al. Presence of diabetes mellitus and steatosis is associated with liver stiffness in a general population: The Rotterdam study. *Hepatology*. 2016;63(1):138-147. doi:10.1002/hep.27981

209. Thwaites PA, Roberts SK, Kitson M KW. Supersonic Shear Imaging and Transient Elastography in Obesity. *Clin Gastroenterol Hepatol*. 2016;14(3):484.

210. Myers RP, Pomier-Layrargues G, Kirsch R, et al. Feasibility and diagnostic performance of the FibroScan XL probe for liver stiffness measurement in overweight and obese patients. *Hepatology*. 2012;55:199-208. doi:10.1002/hep.24624

211. Arena U, Vizzutti F, Abraldes JG, et al. Reliability of transient elastography for the diagnosis of advanced fibrosis in chronic hepatitis C. *Gut*. 2008;57:1288-1293. doi:10.1136/gut.2008.149708

212. Liu C-J. Prevalence and risk factors for non-alcoholic fatty liver disease in Asian people who are not obese. *J Gastroenterol Hepatol*. 2012;27(10):1555-1560. doi:10.1111/j.1440-1746.2012.07222.x

213. Padoin AV, Mottin CC, Moretto M, et al. A comparison of wedge and needle hepatic biopsy in open bariatric surgery. *Obes Surg*. 2006;16:178-182. doi:10.1381/096089206775565159

214. Armstrong MJ, Corbett C, Hodson J, et al. Operator training requirements and diagnostic accuracy of Fibroscan in routine clinical practice. *Postgrad Med J*. 2013;89(1058):685-692. doi:10.1136/postgradmedj-2012-131640

215. Boursier J, Konate A, Guilluy M, et al. Learning curve and interobserver reproducibility evaluation of liver stiffness measurement by transient elastography. *Eur J Gastroenterol Hepatol*. 2008;20(7):693-701. doi:10.1097/MEG.0b013e3282f51992

216. Pang JXQ, Pradhan F, Zimmer S, et al. The feasibility and reliability of transient elastography using Fibroscan: A practice audit of 2335 examinations. *Can J Gastroenterol Hepatol*. 2014;28(3):143-149. http://www.pulsus.com/journals/pdf_frameset.jsp?jnlKy=2&atlKy=12740&isArt=t&jnlAdvert=Gastro&adverifHCTp=&sTitle=The%5Cnfeasibility%5Cnand%5Cnreliability%5Cnof%5Cntransient%5Cnelastography%5Cnusing%5CnFibroscan?:%5CnA%5Cnpractice%5Cnaudit%5Cn2335%5Cnexami.

217. Friedrich-Rust M, Buggisch P, de Knegt RJ, et al. Acoustic radiation force impulse imaging for non-invasive assessment of liver fibrosis in chronic hepatitis B. *J Viral Hepat*. 2013;20(4):240-247. doi:10.1111/j.1365-2893.2012.01646.x

218. Kim SU, Oh HJ, Wanless IR, Lee S, Han KH, Park YN. The Laennec staging system for histological sub-classification of cirrhosis is useful for stratification of prognosis in patients with liver cirrhosis. *J Hepatol*. 2012. doi:10.1016/j.jhep.2012.04.029

219. Ye XP, Ran HT, Cheng J, et al. Liver and spleen stiffness measured by acoustic radiation force impulse elastography for noninvasive assessment of liver fibrosis and esophageal varices in patients with chronic hepatitis B. *J Ultrasound Med*. 2012;31(8):1245-1253. doi:31/8/1245 [pii]

220. Leung HK, Gibson RG, Sood S, Nadebaum D, Gorelik A NA. Liver Stiffness by ARFI does not Correlate with Decompensation and Portal Hypertension in Patients with Cirrhosis. *J Diges Dis Diag*. 2017;1(1):28-37.

221. Schuppan D, Afdhal NH. Seminar Liver cirrhosis. *Lancet*. 2008. doi:10.1016/S0140-6736(08)60383-9

222. Gao J, Ran HT, Ye XP, Zheng YY, Zhang DZ, Wang ZG. The stiffness of the liver and spleen on ARFI Imaging pre and post TIPS placement: A preliminary observation. *Clin Imaging*. 2012;36:135-141. doi:10.1016/j.clinimag.2011.11.014

223. Ye X, Ran H, Cheng J, Zhu Y, Zhang D. Liver and Spleen Stiffness Measured by

Acoustic Radiation Force Impulse Elastography for Noninvasive Assessment. 2012:1245-1253.

224. Colecchia A, Montrone L, Scaioli E, et al. Measurement of spleen stiffness to evaluate portal hypertension and the presence of esophageal varices in patients with HCV-related cirrhosis. *Gastroenterology*. 2012;143(3):646-654. doi:10.1053/j.gastro.2012.05.035

225. Thiele M, Detlefsen S, Sevelsted M??ller L, et al. Transient and 2-Dimensional Shear-Wave Elastography Provide Comparable Assessment of Alcoholic Liver Fibrosis and Cirrhosis. *Gastroenterology*. 2016;150(1):123-133. doi:10.1053/j.gastro.2015.09.040

226. Ferraioli G, Tinelli C, Zicchetti M, et al. Reproducibility of real-time shear wave elastography in the evaluation of liver elasticity. *Eur J Radiol*. 2012;81(11):3102-3106. doi:10.1016/j.ejrad.2012.05.030

227. Sporea I, Bota S, Jurchis A, et al. Acoustic radiation force impulse and supersonic shear imaging versus transient elastography for liver fibrosis assessment. *Ultrasound Med Biol*. 2013;39:1933-1941. doi:10.1016/j.ultrasmedbio.2013.05.003

228. 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*. 2016. doi:10.1055/s-0042-108431

Appendix 1: 2D-SWE Participant Information Sheet

Royal Melbourne Hospital, Department of Radiology

Participant Information Sheet

Royal Melbourne Hospital

Version #1 Date: 27/08/2014

Project*:* Shear Wave Elastography measurement of liver fibrosis. An analysis of variance of results obtained by a single operator, with a view to minimizing number of measurements

Investigators:

Prof Robert Gibson
Professorial Fellow
Dept of Radiology
Royal Melbourne Hospital
9342 7255

A/Prof Amanda Nicoll
Deputy Director
Dept of Gastroenterology & Hepatology
Royal Melbourne Hospital
9342 7470

Dr David Nadebaum
Research Fellow
Depts of Gastroenterology & Radiology
Royal Melbourne Hospital
9342 7470

Project aims / background

The purpose of this research is to test a new ultrasound technique called 'Toshiba Shear Wave Elastography,' which has been developed for the assessment of liver scarring.

Estimating the amount of scarring in the liver is important when determining the severity of a patient's liver disease. This information is relied upon by doctors when making decisions regarding treatment and medical management. A number of machines using 'Shear Wave' technology have now been developed to test for liver scarring. These are all ultrasound-based, and have been proven safe in clinical use.

Your doctor has referred you for one of these tests called ARFI, which itself looks for liver scarring.

Toshiba has developed a new Shear Wave device, which despite sharing similar ultrasound technology to ARFI, uses new advances in data processing. This new technology helps to improve scan performance, and may make 'Toshiba Shear Wave Elastography' more accurate than currently available machines. Whilst the test has shown good accuracy in assessing liver scarring, further testing is required before the device is ready for routine clinical use.

We aim to analyse a number of the test's properties amongst 50 participants at the Royal Melbourne Hospital. In particular, we plan to determine the number of Shear Wave measurements required in each patient, to obtain an accurate liver assessment. This information will be used to develop guidelines for future use of the test, and will help make the device ready for routine clinical use.

What does participating in the project involve?

You have been referred to the radiology department for an ultrasound examination of your liver (an ARFI). We are requesting to use this newer 'Shear Wave' machine as part of your ultrasound scan.

The testing would occur at the end of your routine ultrasound scan, and would take approximately five (5) minutes. The scan is safe, is not painful, and would be very similar to the preceding ultrasound examination.

Your participation in the project will not require any other activities or commitments.

What are the possible benefits?

The accuracy of Shear Wave is still being tested, and the reliability of its measurements has not been confirmed. The Shear Wave readings taken in this study, will therefore not be used to guide your medical care. Your liver assessment will continue unchanged, and will be based on fully-tested technology (such as ARFI). It is therefore unlikely that you will receive any direct benefit from this project.

Your participation will however provide valuable information, which will help to improve our understanding of Shear Wave technology. This research will assist in the development of this newer machine, and may potentially improve the estimation of liver scarring for future patients.

What are the possible risks?

Shear Wave technology has been tested for a number of years, and is considered safe by both Toshiba and the Australian Government. It has no known risks to your health, and should not cause any discomfort.

Participation is voluntary

Participation in any research project is voluntary. If you do not wish to take part you don't have to. If you decide to take part and later change your mind, you are free to withdraw from the project at any stage.

Your decision whether to take part or not to take part, or to take part and then withdraw, will not affect whether you undergo other routine investigations, or your overall medical care.

How will my privacy be maintained?

Shear Wave measurements will be recorded within an electronic database. No additional information regarding you or your medical conditions will be collected for this study. Your information will be recorded together with a 'code number' generated specifically for the study (your name or hospital number will not be used). The information will be recorded in a secure computer file, and will be stored in a locked office. This information will be permanently destroyed seven (7) years after completion of the project. The recorded information will only be accessible by researchers involved in this project, and representatives of the Melbourne Health Human Research Ethics Committee for the purposes of verifying the conduct of this study.

Any information obtained for the purpose of this research that can identify you will be treated as confidential and securely stored. It will be disclosed only with your permission, or as permitted by law. In any publication and/or presentation, information will be provided in such a way that you cannot be identified.

If you have any questions regarding the study, you should feel free to contact one of the researchers listed above.

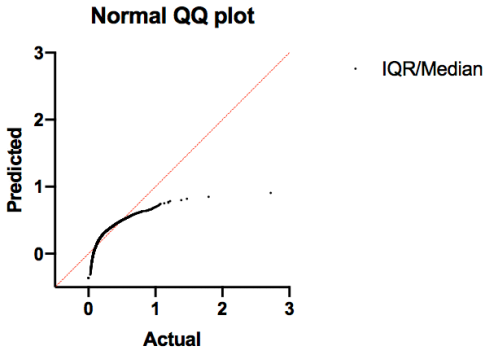
Appendix 2: Statistical test outputs and normality results

ARFI (Siemens)

Normality tests – overall cohort

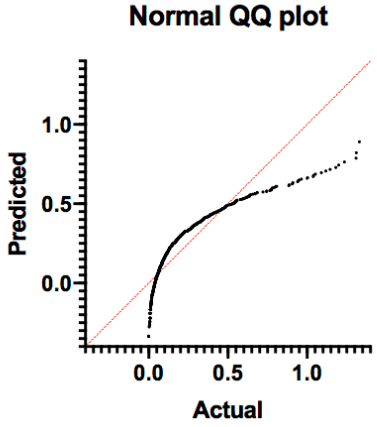
IQR/Median ratios

D'Agostino & Pearson test	
K2	1530
P value	<0.0001
Passed normality test (alpha=0.05)?	No
P value summary	****



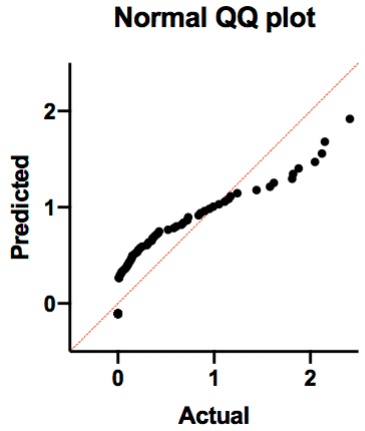
Inter-operator percentage deviation

D'Agostino & Pearson test	
K2	886.2
P value	<0.0001
Passed normality test (alpha=0.05)?	No
P value summary	****



Deviation of ARFI LSMs from biopsy

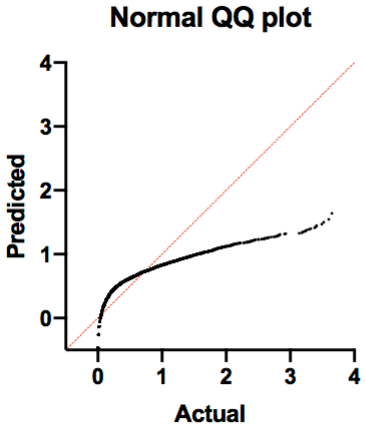
D'Agostino & Pearson test	
K2	41.51
P value	<0.0001
Passed normality test (alpha=0.05)?	No
P value summary	****



N.B. Whilst deviation from biopsy followed a non-Gaussian distribution, parametric testing was still performed (i.e. unpaired t tests) as the mean was felt more representative than the median (due to the high number of measurements which deviated 0 m/s from biopsy). This was felt justifiable given the sample size and weak skew.

Deviation of individual measurements from the set's median

D'Agostino & Pearson test	
K2	13273
P value	<0.0001
Passed normality test (alpha=0.05)?	No
P value summary	****



Chapter 4.41 – Patient demographics

NAFLD Analyses

IQR/Median

Mann Whitney test	
P value	<0.0001
Exact or approximate P value?	Approximate
P value summary	****
Significantly different (P < 0.05)?	Yes
One- or two-tailed P value?	Two-tailed
Sum of ranks in column A,B	1206428 , 550447
Mann-Whitney U	272767
Difference between medians	
Median of column A	0.1537, n=1366
Median of column B	0.2060, n=508
Difference: Actual	0.05230
Difference: Hodges-Lehmann	0.04462

Inter-operator % deviation

Mann Whitney test	
P value	<0.0001
Exact or approximate P value?	Approximate
P value summary	****
Significantly different (P < 0.05)?	Yes
One- or two-tailed P value?	Two-tailed
Sum of ranks in column A,B	300844 , 133003
Mann-Whitney U	70663
Difference between medians	
Median of column A	0.1068, n=678
Median of column B	0.1538, n=253
Difference: Actual	0.04708
Difference: Hodges-Lehmann	0.03761

Inter-operator concordance

P value and statistical significance		
Test	Fisher's exact test	
P value	0.0072	
P value summary	**	
One- or two-sided	Two-sided	
Statistically significant (P < 0.05)?	Yes	
Data analyzed	**NAFLD**	**Non-NAFLD**
Concordant	209	606
Discordant	44	72
Total	253	678
Percentage of row total	**NAFLD**	**Non-NAFLD**
Concordant	25.64%	74.36%
Discordant	37.93%	62.07%
Percentage of column total	**NAFLD**	**Non-NAFLD**
Concordant	82.61%	89.38%
Discordant	17.39%	10.62%

Relationship between NAFLD and BMI

Mann Whitney test	
P value	<0.0001
Exact or approximate P value?	Exact
P value summary	****
Significantly different (P < 0.05)?	Yes
One- or two-tailed P value?	Two-tailed
Sum of ranks in column A,B	13940 , 17185
Mann-Whitney U	3819
Difference between medians	
Median of column A	29.75, n=86
Median of column B	24.93, n=163
Difference: Actual	-4.822
Difference: Hodges-Lehmann	-4.152

Relationship between NAFLD and SLD

Mann Whitney test	
P value	<0.0001
Exact or approximate P value?	Approximate
P value summary	****
Significantly different (P < 0.05)?	Yes
One- or two-tailed P value?	Two-tailed
Sum of ranks in column A,B	629421 , 1129329
Mann-Whitney U	194301
Difference between medians	
Median of column A	2.421, n=508
Median of column B	2.006, n=1367
Difference: Actual	-0.4145
Difference: Hodges-Lehmann	-0.4097

Chapter 4.4.2 Necroinflammatory change

Deviation from biopsy

Unpaired t test	
P value	0.0780
P value summary	ns
Significantly different (P < 0.05)?	No
One- or two-tailed P value?	Two-tailed
t, df	t=1.780, df=105
How big is the difference?	
Mean of column A	0.3255
Mean of column B	0.5389
Difference between means (B - A)	0.2134 ± 0.1199
95% confidence interval	-0.02435 to 0.4511
R squared (eta squared)	0.02928

Chapter 4.4.3 Hepatosteatosis

IQR/Median

Kruskal-Wallis test	
P value	<0.0001
Exact or approximate P value?	Approximate
P value summary	****
Do the medians vary signif. (P < 0.05)?	Yes
Number of groups	3
Kruskal-Wallis statistic	51.81

Inter-operator %deviation

Kruskal-Wallis test	
P value	<0.0001
Exact or approximate P value?	Approximate
P value summary	****
Do the medians vary signif. (P < 0.05)?	Yes
Number of groups	3
Kruskal-Wallis statistic	20.57

Inter-operator concordance

P value and statistical significance	
Test	Chi-square
Chi-square, df	6.956, 2
P value	0.0309
P value summary	*
One- or two-sided	NA
Statistically significant (P < 0.05)?	Yes

Body mass index

IQR/Median

Kruskal-Wallis test	
P value	<0.0001
Exact or approximate P value?	Approximate
P value summary	****
Do the medians vary signif. (P < 0.05)?	Yes
Number of groups	3
Kruskal-Wallis statistic	51.33

Inter-operator % deviation

Kruskal-Wallis test	
P value	<0.0001
Exact or approximate P value?	Approximate
P value summary	****
Do the medians vary signif. (P < 0.05)?	Yes
Number of groups	3
Kruskal-Wallis statistic	26.88

Inter-operator concordance

P value and statistical significance	
Test	Chi-square
Chi-square, df	8.535, 2
P value	0.0140
P value summary	*
One- or two-sided	NA
Statistically significant (P < 0.05)?	Yes

Chapter 4.4.5 SLD

Failure of IQR/Median Criteria

P value and statistical significance	
Test	Chi-square
Chi-square, df	129.9, 2
P value	<0.0001
P value summary	****
One- or two-sided	NA
Statistically significant (P < 0.05)?	Yes

Inter-operator Concordance Rates

P value and statistical significance	
Test	Chi-square
Chi-square, df	41.48, 2
P value	<0.0001
P value summary	****
One- or two-sided	NA
Statistically significant (P < 0.05)?	Yes

Correlation with biopsy (SLD < 2.5cm vs >2.5cm)

P value and statistical significance			
Test	Fisher's exact test		
P value	<0.0001		
P value summary	****		
One- or two-sided	Two-sided		
Statistically significant (P < 0.05)?	Yes		
Data analyzed	**Concordant**	**Discordant**	**Total**
<2.5	70	12	82
>2.5	10	15	25
Total	80	27	107
Percentage of row total	**Concordant**	**Discordant**	
<2.5	85.37%	14.63%	
>2.5	40.00%	60.00%	

Chapter 4.5.1 Measurement Depth

LCD and deviation from set's median

Unpaired t test	
P value	<0.0001
P value summary	****
Significantly different (P < 0.05)?	Yes
One- or two-tailed P value?	Two-tailed
t, df	t=10.26, df=18718
How big is the difference?	
Mean of column A	0.3117
Mean of column B	0.2250
Difference between means (B - A)	-0.08675 ± 0.008452
95% confidence interval	-0.1033 to -0.07018
R squared (eta squared)	0.005597

Percentage deviation from set's median

Unpaired t test	
P value	<0.0001
P value summary	****
Significantly different (P < 0.05)?	Yes
One- or two-tailed P value?	Two-tailed
t, df	t=4.410, df=18718
How big is the difference?	
Mean of column A	0.1521
Mean of column B	0.1304
Difference between means (B - A)	-0.02171 ± 0.004924
95% confidence interval	-0.03137 to -0.01206
R squared (eta squared)	0.001038

LCD and deviation from biopsy

Unpaired t test	
P value	<0.0001
P value summary	****
Significantly different (P < 0.05)?	Yes
One- or two-tailed P value?	Two-tailed
t, df	t=4.349, df=1013
How big is the difference?	
Mean of column A	0.6551
Mean of column B	0.4088
Difference between means (B - A)	-0.2462 ± 0.05662
95% confidence interval	-0.3573 to -0.1351
R squared (eta squared)	0.01833

Chapter 4.5.2 Individual operator experience

Deviation from biopsy

Unpaired t test	
P value	0.0224
P value summary	*
Significantly different (P < 0.05)?	Yes
One- or two-tailed P value?	Two-tailed
t, df	t=2.319, df=104
How big is the difference?	
Mean of column A	0.5877
Mean of column B	0.3011
Difference between means (B - A)	-0.2867 ± 0.1236
95% confidence interval	-0.5318 to -0.04150
R squared (eta squared)	0.04916

Concordance with biopsy

P value and statistical significance			
Test	Fisher's exact test		
P value	0.0701		
P value summary	ns		
One- or two-sided	Two-sided		
Statistically significant (P < 0.05)?	No		
Data analyzed	**Concordant**	**Discordant**	**Total**
<25	16	15	31
>25	54	21	75
Total	70	36	106
Percentage of row total	**Concordant**	**Discordant**	
<25	51.61%	48.39%	
>25	72.00%	28.00%	

Deviation from biopsy (operators >100 scans only)

Unpaired t test	
P value	0.0237
P value summary	*
Significantly different (P < 0.05)?	Yes
One- or two-tailed P value?	Two-tailed
t, df	t=2.312, df=70
How big is the difference?	
Mean of column A	0.5746
Mean of column B	0.2604
Difference between means (B - A)	-0.3142 ± 0.1359
95% confidence interval	-0.5852 to -0.04314
R squared (eta squared)	0.07094

4.5.3. Institutional experience

IQR/Median

Mann Whitney test	
P value	0.0338
Exact or approximate P value?	Approximate
P value summary	*
Significantly different (P < 0.05)?	Yes
One- or two-tailed P value?	Two-tailed
Sum of ranks in column A,B	289857 , 1467018
Mann-Whitney U	211698

Inter-operator deviation

Mann Whitney test	
P value	0.0013
Exact or approximate P value?	Approximate
P value summary	**
Significantly different (P < 0.05)?	Yes
One- or two-tailed P value?	Two-tailed
Sum of ranks in column A,B	75135 , 358712
Mann-Whitney U	46267
Difference between medians	
Median of column A	0.1622, n=141
Median of column B	0.1101, n=790
Difference: Actual	-0.05204
Difference: Hodges-Lehmann	-0.03457

Inter-operator concordance

P value and statistical significance			
Test	Fisher's exact test		
P value	0.1392		
P value summary	ns		
One- or two-sided	Two-sided		
Statistically significant (P < 0.05)?	No		
Data analyzed	**Concordant**	**Discordant**	**Total**
<150 scans	119	24	143
>150 scans	692	97	789
Total	811	121	932
Percentage of row total	**Concordant**	**Discordant**	
<150 scans	83.22%	16.78%	
>150 scans	87.71%	12.29%	

4.6.2. SLD Criteria

Biopsy concordance for patients passing SLD and IQR/Median criteria vs those failing either.

P value and statistical significance	
Test	Fisher's exact test
P value	0.0012
P value summary	**
One- or two-sided	Two-sided
Statistically significant (P < 0.05)?	Yes

4.6.3. Inter-operator concordance

Deviation from biopsy depending on % inter-operator deviation

Unpaired t test	
P value	0.0025
P value summary	**
Significantly different (P < 0.05)?	Yes
One- or two-tailed P value?	Two-tailed
t, df	t=3.117, df=86
How big is the difference?	
Mean of column A	0.1876
Mean of column B	0.5998
Difference between means (B - A)	0.4122 ± 0.1322
95% confidence interval	0.1493 to 0.6751
R squared (eta squared)	0.1015

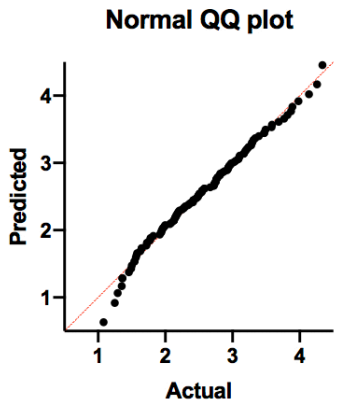
Deviation from biopsy amongst higher vs. lower operator values

Unpaired t test	
P value	0.0925
P value summary	ns
Significantly different (P < 0.05)?	No
One- or two-tailed P value?	Two-tailed
t, df	t=1.777, df=18
How big is the difference?	
Mean of column A	0.1670
Mean of column B	0.4270
Difference between means (B - A)	0.2600 ± 0.1463
95% confidence interval	-0.04744 to 0.5674
R squared (eta squared)	0.1492

4.7 Cirrhosis Assessment.

Distribution – ARFI LSM

D'Agostino & Pearson test	
K2	2.762
P value	0.2513
Passed normality test (alpha=0.05)?	Yes
P value summary	ns



ARFI LSM in Child Pugh A vs. B/C

Unpaired t test	
P value	0.0005
P value summary	***
Significantly different (P < 0.05)?	Yes
One- or two-tailed P value?	Two-tailed
t, df	t=3.576, df=120
How big is the difference?	
Mean of column A	2.360
Mean of column B	2.816
Difference between means (B - A)	0.4559 ± 0.1275
95% confidence interval	0.2035 to 0.7084
R squared (eta squared)	0.09630

Portal hypertension

Unpaired t test	
P value	0.0411
P value summary	*
Significantly different (P < 0.05)?	Yes
One- or two-tailed P value?	Two-tailed
t, df	t=2.051, df=339
How big is the difference?	
Mean of column A	2.612
Mean of column B	2.447
Difference between means (B - A)	-0.1643 ± 0.08010
95% confidence interval	-0.3218 to -0.006692
R squared (eta squared)	0.01225

Hepatic encephalopathy

Unpaired t test	
P value	0.0312
P value summary	*
Significantly different (P < 0.05)?	Yes
One- or two-tailed P value?	Two-tailed
t, df	t=2.164, df=345
How big is the difference?	
Mean of column A	2.493
Mean of column B	2.758
Difference between means (B - A)	0.2649 ± 0.1225
95% confidence interval	0.02408 to 0.5058
R squared (eta squared)	0.01339

Ascites

Unpaired t test	
P value	0.0360
P value summary	*
Significantly different (P < 0.05)?	Yes
One- or two-tailed P value?	Two-tailed
t, df	t=2.105, df=339
How big is the difference?	
Mean of column A	2.486
Mean of column B	2.699
Difference between means (B - A)	0.2125 ± 0.1009
95% confidence interval	0.01396 to 0.4110
R squared (eta squared)	0.01291

Oesophageal Varices

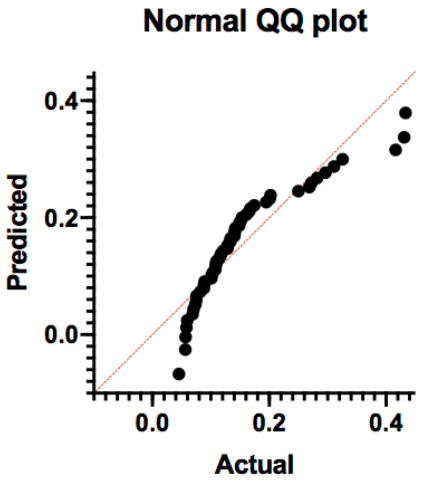
Unpaired t test	
P value	0.0038
P value summary	**
Significantly different (P < 0.05)?	Yes
One- or two-tailed P value?	Two-tailed
t, df	t=2.954, df=113
How big is the difference?	
Mean of column A	2.299
Mean of column B	2.671
Difference between means (B - A)	0.3722 ± 0.1260
95% confidence interval	0.1226 to 0.6218
R squared (eta squared)	0.07170

Toshiba 2D-SWE

Chapter 6.5

IQR/Median - distribution

D'Agostino & Pearson test	
K2	20.41
P value	<0.0001
Passed normality test (alpha=0.05)?	No
P value summary	****



IQR/Median and BMI

Mann Whitney test	
P value	0.0106
Exact or approximate P value?	Exact
P value summary	*
Significantly different (P < 0.05)?	Yes
One- or two-tailed P value?	Two-tailed
Sum of ranks in column A,B	415 , 1125
Mann-Whitney U	205
Difference between medians	
Median of column A	0.1118, n=20
Median of column B	0.1486, n=35
Difference: Actual	0.03673
Difference: Hodges-Lehmann	0.04278

IQR/Median and hepatosteatosis

Mann Whitney test	
P value	0.3142
Exact or approximate P value?	Exact
P value summary	ns
Significantly different (P < 0.05)?	No
One- or two-tailed P value?	Two-tailed
Sum of ranks in column A,B	1095 , 445
Mann-Whitney U	234
Difference between medians	
Median of column A	0.1212, n=41
Median of column B	0.1408, n=14
Difference: Actual	0.01962
Difference: Hodges-Lehmann	0.02362

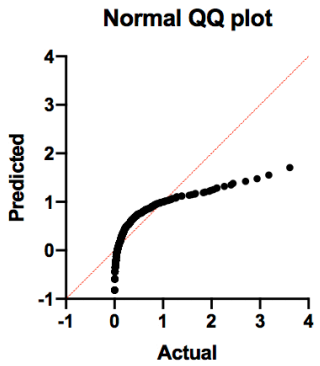
IQR/Median and NAFLD

Mann Whitney test	
P value	0.1489
Exact or approximate P value?	Exact
P value summary	ns
Significantly different (P < 0.05)?	No
One- or two-tailed P value?	Two-tailed
Sum of ranks in column A,B	556 , 984
Mann-Whitney U	243
Difference between medians	
Median of column A	0.1415, n=17
Median of column B	0.1172, n=38
Difference: Actual	-0.02426
Difference: Hodges-Lehmann	-0.02935

Deviation of individual measurements from the set's median

Distribution

D'Agostino & Pearson test	
K2	430.6
P value	<0.0001
Passed normality test (alpha=0.05)?	No
P value summary	****



Measurements deviation – LCD <1.5cm vs. >1.5cm

Unpaired t test	
P value	<0.0001
P value summary	****
Significantly different (P < 0.05)?	Yes
One- or two-tailed P value?	Two-tailed
t, df	t=4.316, df=546
How big is the difference?	
Mean of column A	0.3886
Mean of column B	0.2248
Difference between means (B - A)	-0.1638 ± 0.03795
95% confidence interval	-0.2383 to -0.08923
R squared (eta squared)	0.03299

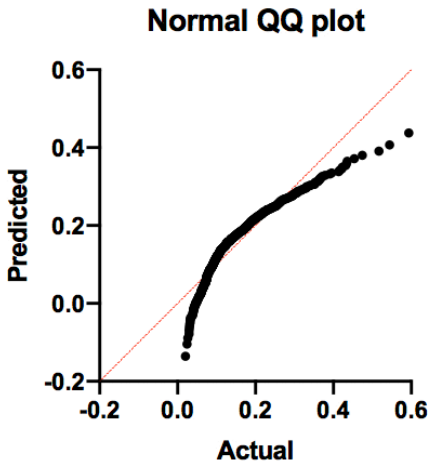
Measurement deviation - ROI SD/Speed <0.15 vs. >0.15

Unpaired t test	
P value	<0.0001
P value summary	****
Significantly different (P < 0.05)?	Yes
One- or two-tailed P value?	Two-tailed
t, df	t=5.324, df=545
How big is the difference?	
Mean of column A	0.2186
Mean of column B	0.4214
Difference between means (B - A)	0.2028 ± 0.03810
95% confidence interval	0.1280 to 0.2777
R squared (eta squared)	0.04944

ROI SD / Speed

Distribution

D'Agostino & Pearson test	
K2	140.1
P value	<0.0001
Passed normality test (alpha=0.05)?	No
P value summary	****



ROI SD/Speed and LCD <1cm vs. >1cm

Mann Whitney test	
P value	<0.0001
Exact or approximate P value?	Exact
P value summary	****
Significantly different (P < 0.05)?	Yes
One- or two-tailed P value?	Two-tailed
Sum of ranks in column A,B	33093 , 116786
Mann-Whitney U	14860
Difference between medians	
Median of column A	0.1694, n=96
Median of column B	0.1173, n=451
Difference: Actual	-0.05213
Difference: Hodges-Lehmann	-0.03995