Biomarker Qualification Letter of Intent (LOI) Content Elements

1. Administrative Information

1. Submission Title

Imaging Biomarkers for Diagnosis of Nonalcoholic Steatohepatitis (NASH)

2. Requesting Organization

Name of the organization:

Foundation for the National Institutes of Health Biomarkers Consortium 11400 Rockville Pike, Suite 600 North Bethesda, MD 20852
Phone: (301) 402-5311
Website: https://fnih.org/what-we-do/biomarkers-consortium

Primary point of contact:

Tania Kamphaus, PhD
Scientific Program Manager
Phone: (301) 435-6247
Email: TKamphaus@fnih.org

Alternate point of contact:

Anthony E. Samir, MD, MPH (asamir@mgh.harvard.edu)
Claude B. Sirlin, MD (csirlin@health.ucsd.edu)
Sarah Sherlock, PhD (sarah.sherlock@pfizer.com)

Supporting or participating organizations or individuals:

FNIH Biomarkers Consortium: **Non-Invasive** Biomarkers of Meta**B**olic **L**iver Diseas**E** (NIMBLE)

Project Steering Committee:
Arun Sanyal, MD (VCU)
Rohit Loomba, MD (UCSD)
Claude B. Sirlin, MD (UCSD)
Anthony E. Samir, MD (MGH)
Sudha Shankar, MD (AstraZeneca)
Roberto Calle, MD (Pfizer)
Sarah Sherlock, PhD (Pfizer)

3. Submission Date(s)

LOI submission date (original): December 10, 2020

PREAMBLE

The context of use (COU) proposed in this LOI is consistent with the designated goals of the NIMBLE program, may apply to biomarkers in combination, and is not intended to limit or suggest a refinement of any of the individual biomarkers outlined below.

2. Introduction

Nonalcoholic fatty liver disease (NAFLD) affects 30% of adults in the United States and is characterized by accumulation of excess triglycerides in the liver (Younossi et al. 2016). By definition, it is not due to consumption of harmful amounts of alcohol. It has two principal subtypes: 1) nonalcoholic fatty liver (NAFL), and 2) nonalcoholic steatohepatitis (NASH). NASH, especially in those who have developed some fibrosis, progresses to cirrhosis and end-stage liver disease more frequently than NAFL. The identification of patients at risk of progression to cirrhosis and clinically meaningful adverse outcomes and the development of effective therapies for these patients is a public health priority. Both NAFL and NASH are asymptomatic until advanced disease is present and cannot be readily diagnosed or distinguished from each other by clinical and routine laboratory tests. Diagnosis and staging of NASH therefore currently rely on liver biopsy with histologic assessment, which comprises the current reference standard for this purpose. Liver biopsies are invasive, painful and carry a small risk of catastrophic complications including death. Furthermore, histologic evaluation using liver biopsy is limited by moderate accuracy and by sampling variability, due to the small amount of material gathered for examination relative to the size of the liver, and by interpreter variability.

Akin to the situation in clinical practice described above where diagnosis of NASH depends on histologic assessment, the main way to identify patients with NASH for participation in drug development trials is liver biopsy. Patients are selected to undergo screening with liver biopsy by assessing their probability of having NASH, which depends on clinical history and non-specific laboratory values. Unfortunately, current selection of subjects for biopsy via clinical means is imperfect and leads to a high frequency of biopsies in patients that end up not meeting histopathologic entry criteria for NASH necessary for inclusion in drug development trials. The screen-failure rates across Phase 2b-3 trials for NASH range from 50 to 80% (Ampuero and Romero-Gomez 2020). *The lack of reliable non-invasive means of enriching the population of subjects undergoing liver biopsy likely to meet the histopathologic criteria for NASH with fibrosis represents a major barrier for identifying subjects eligible for pharmacologic intervention and therefore for enrollment in drug development trials. The development and qualification of such markers would expedite drug development trials and accelerate the development of pharmacologic therapeutic interventions for NASH. These are patients that on biopsy show evidence of active NASH (defined by presence of steatohepatitis and a NAFLD activity score (NAS) of 4 or greater) with fibrosis score ≥ 2, and are thus identified as a population of patients with NASH with a higher probability to progress to adverse liver outcomes due to the stage of fibrosis.*

This underscores the urgent need for robust, reliable, readily deployable non-invasive tools that identify subjects with a high probability of meeting histopathologic criteria for active NASH with fibrosis stage 2, or greater. In considering those with NASH and fibrosis stage 2 or greater, it is also important to distinguish those with cirrhosis (stage 4 fibrosis) from those with earlier stage disease because the disease biology and clinical course differ, requiring different approaches to therapy and assessment. In this LOI, we focus on those with NASH with high disease activity (NAS ≥ 4) and fibrosis stages 2-3 that we refer to as 'at risk' NASH.

This application is limited to testing seven selected ultrasound- and magnetic resonance imaging (MRI)-based biomarkers which may eventually be considered as stand-alone biomarkers or as part of a panel of imaging biomarkers in combination, to identify active NASH (NAS ≥ 4) with stage 2-3 fibrosis (i.e., 'at risk' NASH).

Summary data for each of the biomarkers is provided in **Table 2.1**.

Table 2.1. NIMBLE ultrasound- and MRI-based biomarkers

Biomarker	NASH	Steatosis	Fibrosis
Ultrasound			
SWE-based SWS	X		X
VCTE-based E	X		X
VCTE-based CAP		X	
MRI			
MRI-based PDFF		X	
MRI-based cT1	X		X
2D MRE-based liver stiffness	X		X
3D MRE-based liver stiffness	X		X

SWE = shear wave elastography; SWS = shear wave speed; VCTE – vibration-controlled transient elastography; E= Young's modulus; CAP = controlled attenuation parameter; MRI = magnetic resonance imaging; PDFF = proton density fat fraction; cT1 = Iron corrected T1 relaxation time; MRE = magnetic resonance elastography

The FNIH NIMBLE team selected these imaging biomarkers based on publicly available data (peer-reviewed publications) supporting their analytical and clinical performance for assessment of inflammation (NASH), steatosis, fibrosis, or a combination of the above – histologic features relevant to the proposed context of use. Supplementary data summaries provided by biomarker developers were considered. In addition to biological plausibility and emerging clinical evidence, additional attractive features of ultrasound- and MRI-based biomarkers are that they can be acquired safely, without contrast agents or ionizing radiation. **This LOI describes a complementary effort to a prior LOI submitted by the NIMBLE project (DDTBMQ000084) for circulating biomarkers addressing the same context of use.**

It is possible that some biomarkers are superior in performance to others. Decisions on the final biomarker, or biomarker panel (to be part of a future qualification plan submission) will be made after all initially selected biomarkers described in the two NIMBLE LOIs are tested and analyzed.

Supporting references are available in **Appendix 8**.

The drug development need that will be addressed by this qualification is diagnostic enrichment of 'at-risk' NASH patients for inclusion in drug development trials.

The development of NASH follows a complex interplay of mechanisms where genetic susceptibility coupled with environmental factors such as sedentary lifestyle and excess dietary intake results in hepatic steatosis, insulin resistance, enhanced *de novo* lipogenesis, lipotoxicity, hepatocyte oxidative stress and mitochondrial dysfunction, and activation of inflammatory pathways with inflammatory infiltration, ultimately progressing, in some patients to fibrosis and organ damage, including cirrhosis. The multifaceted nature of the pathogenesis of NASH means that the influence of multiple contributory factors needs to be taken into account and quantified to enable non-invasive assessment of risk of individual patients to progress to adverse liver outcomes. This is true whether we are trying to show a direct association between biomarkers and clinical outcomes, or through an intermediate step such as showing an association with histologic liver findings (e.g., fibrosis) that are themselves associated with liver outcomes. A number of circulating and imaging-based biomarkers have been shown to be associated with liver biopsy histologic features. These biomarkers provide read-outs on some of these varied pathways contributing to NASH pathogenesis. However, these biomarkers have not been systematically and independently evaluated. The NIMBLE Project's overarching objective is, through a systematic evaluation of these promising biomarkers (both individually and in combination) to identify a biomarker (or biomarker panel) that will be both robust and sensitive to identify 'at-risk' NASH patients.

3. Context of Use Statement *(all candidate biomarkers)*

A noninvasive imaging-based diagnostic enrichment biomarker intended for use, in conjunction with clinical factors and/or circulating biomarkers, to identify patients likely to have liver histopathologic findings of nonalcoholic steatohepatitis (NASH) and with a nonalcoholic fatty liver disease activity score (NAS) ≥ 4 and liver fibrosis stages 2 or 3 (by Brunt/Kleiner scale); and thus appropriate for inclusion in liver biopsy-based NASH drug development clinical trials focused on pre-cirrhotic stages of NASH.

4. General Considerations *(all candidate biomarkers)*

4.1 Performance measures

This LOI focuses on measures of accuracy using histology as the reference standard, and on measures of precision.

The main measures of accuracy are AUROC and diagnostic performance at cutoffs. With regard to cutoffs, Mehta et al (2009) concluded that using histologic assessments derived from liver biopsy as reference values in typical diagnostic accuracy studies (i.e., those assuming cutoffs based on 90% sensitivity, 90% specificity, and 40% significant disease prevalence), the best possible AUROC is about 0.90 for a perfect biomarker. This best possible value should be kept in mind when interpreting the reported AUROCs for the biomarkers being discussed.

The main measures of precision, with their accompanying definitions are as follows (Yokoo et al. 2018):

- Repeatability: agreement between repeated measurements under identical or near-identical conditions (e.g., scan-rescan measurements without changing acquisition method or scanning parameters, and
- Reproducibility: agreement between repeated measurements under different conditions (e.g., using different equipment, different scanning parameters, different acquisition or analysis methods, or different readers/analysts).

4.2 Quality assurance and quality control

This LOI describes the quality assurance (QA) and quality control (QC) procedures relevant for each biomarker. These terms are as follows:

- QA refers to procedures to ensure that the process of measuring a particular biomarker (including imaging equipment, acquisition, and analysis) is adequate.
- QC refers to procedures to inspect and verify the quality of the images acquired and measurements made.

4.3 Terminology

As explained in **Sections 5.1** and **5.2**, ultrasound-based SWE techniques (point SWE, 2D SWE, VCTE) measure shear wave speed (*SWS*), which has units of m/s. *SWS* can be converted to the Young's modulus (*E*), which has units of kPa, using the following formula:

$E = 3\rho(SWS)^2$, where E represents Young's modulus, ρ represents tissue density, and *SWS* is shear wave speed.

For ultrasound-based SWE, although both measures (*SWS* and *E*) have been used in the literature, SWS is preferred, and manufacturers' guidelines and the QIBA SWE profile recommend reporting the SWS rather than *E*. The reason is that *SWS* is what is actually measured while *E* is derived from *SWS* under simplifying assumptions. Accordingly, in this LOI, *SWS* is used preferentially rather than *E* unless the cited study(ies) report only *E*. For VCTE, *E* is preferred, as this is used in the existing literature and reported by default on Fibroscan devices.

4.4 Cutoffs

This LOI includes published diagnostic cutoff values for histologic classification by each candidate biomarker. These cutoff values are provided as examples. They may not represent the exact cutoffs to be proposed for the specified context of use (**Section 3**).

5. Ultrasound Biomarker Information

This section contains information on biomarker interpretation and information, analytical considerations, and supporting information for each of the following three categories of ultrasound-based candidate biomarkers:

- Shear wave elastography- (SWE-) based shear wave speed (*SWS*)
- Vibration-controlled transient elastography- (VCTE-) based liver stiffness
- Vibration-controlled transient elastography- (VCTE-) based controlled attenuation parameter (CAP)

SWE SWS

5.1 SWE-based SWS

5.1.1 Biomarker interpretation information

5.1.1.1 Biomarker name

Shear-wave elastography- (SWE-) based shear wave speed (*SWS*)

5.1.1.2 Biomarker type

Diagnostic enrichment biomarker, adapted from the BEST Glossary

5.1.1.3 Analytical methods

SWE-based *SWS* is a non-invasive biomarker of hepatic fibrosis in NAFLD (Ozturk et al. 2018a) and can be measured using proprietary techniques available on clinical ultrasound systems.

These techniques dynamically stress the liver using one or more acoustic radiation force impulse (ARFI) pulses to generate parallel or perpendicular acoustic shear waves within the liver. There are currently two classes of ultrasound-based shear wave imaging techniques: a) Point Shear Wave Elastography (pSWE), which measures *SWS* at a single focal location called a region of interest (ROI), and b) 2D Shear Wave Elastography (SWE), which measures *SWS* in a 2D region and displays the data as a spatial distribution map called an 'elastogram'. An ROI is placed within the elastogram and the mean *SWS* values from that ROI are recorded. ROIs and elastograms are user defined and overlain on anatomic B-mode images. Both classes of SWE can be performed on conventional US machines using standard ultrasound probes. These methods are summarized in **Table 5.1**.

5.1.1.4 Measurement units and limit(s) of detection

Shear wave speed *(SWS)* values are expressed in units of meters per second (m/s). As described earlier, SWE results can also be reported as the **Young's modulus** (*E*), the units of which are expressed in units of kilopascals (kPa) (Ozturk et al. 2018a).

Limits of detection for SWE-based *SWS* and *E* can range from ~ 0 to 4 m/s, and from ~ 0 to 120 kPa, respectively, depending on ultrasound system settings and ultrasound scanner manufacturer.

Table 5.1. 2D-SWE and pSWE technique description*

	pSWE	2D-SWE
Excitation method	Dynamic stress generated by an ARFI pulse applied at a single, focal zone	Dynamic stress generated by ARFI pulses applied in multiple focal zones interrogated in rapid succession, faster than the shear wave speed, creating a near-cylindrical shear wave cone.
Measurables	Shear wave speed reported directly, sometimes converted to Young's modulus, in the focal zone	Real-time generation of an elastogram displaying the distribution of shear wave speed or Young's modulus.
ROI selection	Operator uses B-mode ultrasound (US) to select ROI location	Operator uses B-mode ultrasound (US) to select elastogram location. Operator then selects ROI within a homogeneous or representative portion of elastogram
Images generated	ROI overlay on B-mode image, but no SWE colorized overlay generated	Quantitative elastogram generated. Elastogram and user-selected ROI within elastogram superimposed on B-mode images

** - adapted from* (Sigrist et al. 2017)

5.1.1.5 Biomarker interpretation and utility

Liver SWE-based *SWS* is interpreted and used directly, without additional processing or modification.

Table 5.2, based on a study showing that SWE-based *SWS* values correlate with hepatic histologic fibrosis stage reference values (Takeuchi et al. 2018), illustrates covariation between the biomarker and fibrosis stage. The actual cutoff values to be validated may differ from those listed in **Table 5.2**.

Notice that the AUROC values in **Table 5.2** are close to the expected values for a perfect biomarker under typical study conditions as explained in **Section 4.1**.

A common **actionable conclusion** (use case) for SWE-based *SWS* as a diagnostic enrichment biomarker in drug development trials is to enrich a population sample at screening for histologic fibrosis stage above a certain histologic fibrosis stage. For example, a SWE-based *SWS* cutoff at or near 11.5 kPa could be used to enrich a trial intending to enroll subjects with histologic fibrosis stages ≥ 2. We anticipate that a higher cutoff could be used to exclude subjects likely to have cirrhosis; an appropriate cutoff for this purpose has not yet been identified and will be assessed in NIMBLE studies.

5.1.2 Analytical considerations

5.1.2.1 Method of measurement

SWE-based *SWS* is acquired and calculated as described in **Section 5.1.1.3** (Sigrist et al. 2017).

5.1.2.2 Quality assurance and quality control

QA and **QC** procedures are described in the Quantitative Imaging Biomarkers Alliance (QIBA) Profile for liver SWS (Hall et al. 2013; QIBA).

Key components are summarized below.

Figure 5.1. **a)** Transducer face is parallel to liver capsule, and SWE 'box' is located at the center of the image. **b)** When transducer face is not positioned parallel to liver capsule, ARFI pulse power loss may occur and signal quality may be deteriorated, which will result in unreliable SWE measurements.

Single vendor image for illustrative purposes only

Patient positioning: Patients should be positioned comfortably and instructed to stay still during acquisitions.

Patient breath-holding: breath holding or suspended tidal respiration is suggested for all acquisitions (QIBA). Subjects should not take a deep breath during acquisitions, as diaphragm motion may mask the liver.

Table 5.2. SWE-based cutoff values (Young's modulus) by histologic fibrosis stage

Histologic classification	**AUROC** (with 95% CI)	**Cutoff value** (kPa)	**Sensitivity** (%)	**Specificity** (%)
Any fibrosis (≥ Stage 1)	0.82 (0.66-0.91)	6.6	79	67
Significant fibrosis (≥ Stage 2)	0.75 (0.62-0.85)	11.6	52	44
Advanced fibrosis (≥ Stage 3)	0.82 (0.70-0.90)	13.1	63	57
Cirrhosis (Stage 4)	0.90 (0.78-0.96)	15.7	100	82

Data in this table from (Takeuchi et al. 2018)*, which reported only Young's modulus*

Transducer positioning: The transducer face should be positioned parallel to the liver capsule (**Figure 5.1a**). Otherwise, the ARFI pulse(s) may be weak, which may result in inadequate signal transmission and, for 2D-SWE variants, incomplete elastogram filling (**Figure 5.1b**).

ROI and elastogram placement: ROIs and elastogram images need to be positioned on the center-line of the image (**Figure 5.2**) and away from vessels, gallbladder, liver capsule, large bile ducts, and any incidental lesions. Positioning the elastogram and ROI away from these structures is important as SWE evaluation algorithms assume homogeneous isotropic tissue (liver parenchyma), not heterogeneous tissue (vessels, gallbladder, etc.) or lesions. Positioning elastograms and ROIs away

from the image center-line may introduce variations in *SWS* results. ROIs need to be positioned in the most homogenous region of the elastogram color map.

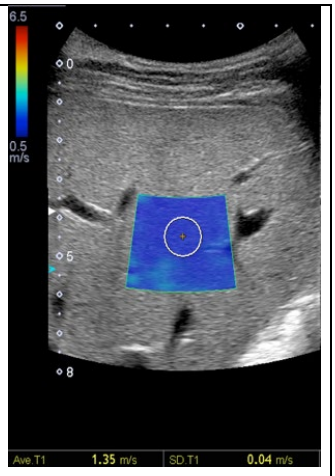
ROI size: Although the effect of ROI size has not been systematically examined, in the QIBA SWS profile it is recommended that ROIs be at least 1 cm diameter (QIBA).

Verification: B-mode images should be obtained immediately before and after *SWS* acquisition to confirm lack of interim liver movement.

Measurement number: There is variation in the number of recommended *SWS* measurements across different manufacturers and different guidelines. For the purposes of this LOI, we have selected a conservative (higher) number of 10 required measurements. Operators are instructed to maximize measurement quality through the techniques described above. If measurement quality is deemed suboptimal by the operator, manifesting as noticeable motion artifact or, where applicable, incomplete filling of the region of interest with colored pixels, operators may repeat the measurement procedure up to twenty times and then select the 10 best measurements. A total of 20 attempted measurements was selected based on expert opinion regarding the number of measurements patients would tolerate and that could be completed in a clinically reasonable time period.

Figure 5.2. High quality image SWE example. Elastogram and ROI located at image mid-line.

- Transducer face parallel to liver capsule
- Elastogram located in 'sweet spot area' or area between 2 specific points (2 cm from liver capsule, 6.5 cm from skin)
- Elastogram located in an area away from vessels, gallbladder, or any confounding structure.
- ROI adjusted to 1 cm diameter.



Single vendor image for illustrative purposes only.

Quantitative QC metrics. As described in the Society of Radiologist in Ultrasound Liver Elastography Consensus Statement (Barr et al. 2020), the interquartile range divided by the median *SWS* or estimated Young's modulus (*E*) may be used as a quality metric. For *SWS* measures this should be less than 15%. For *E* measures this should be less than 30%. The quality metric varies across parameters as the relationship between *SWS* and *E* is not linear.

5.1.2.3 Validation studies to date

The **measurement tool** is the standard *SWS* exam (**Section 5.1.1.3**).

No explicit **device calibrations** are required for *SWS* measurements apart from the standard general-purpose ultrasound calibrations required for all ultrasound scanners. Thus, validation studies to date have not required device calibration.

Validation studies have been conducted to assess diagnostic accuracy (Ozturk et al. 2020), and measures of precision, including same-operator test-retest repeatability, different-operator reproducibility (Hudson et al. 2013; Mancini et al. 2017), and different-device reproducibility (Lee et al. 2019).

Also, as part of the NIH subaward HHSN268201500021C, same-operator test-retest repeatability and different-operator reproducibility under QIBA profile suggestions, were reported to the RSNA QIBA leadership. The manuscript is under production. The unpublished results are provided in **Appendices 1** and **2**.

5.1.2.4 Final biomarker validation

NIMBLE studies will provide more comprehensive assessments of *SWS*-based liver stiffness precision and, using histology as the reference standard, accuracy.

5.1.2.5 Additional considerations

Hepatic SWE-based *SWS* acquisition and analysis have been standardized in published validation studies (Dhyani et al. 2017; Joo et al. 2019), drug development trials (**Appendix 3**), and as per the QIBA SWS Profile (QIBA).

	5.1.3 Supporting information **5.1.3.1 Underlying biological process** SWE-based *SWS* is a quantitative imaging biomarker of hepatic fibrosis (Sigrist et al. 2017), and hepatic fibrosis is one of the **underlying biological processes** contributing to the development and progression of NASH (Ozturk et al. 2020). **5.1.3.2 Clinical data to support the biomarker in its COU** SWE-based *SWS* has been used in numerous NASH drug development trials as a diagnostic enrichment biomarker, summarized in **Appendix 3**, from ***clinicaltrials.gov***. **5.1.3.3 Planned studies to support the biomarker and COU** The NIMBLE studies will address the gaps in knowledge for this COU described above in **Section 7.5**. **5.1.3.4 Alternative comparator, current standard(s), or approaches** **Alternate comparator imaging biomarkers** to assess hepatic fibrosis include MRI-based biomarkers of hepatic fibrosis, which are being assessed in NIMBLE studies (see **Section 6**).
VCTE-based E, and equiv. SWS	**5.2 VCTE-based *E* and equivalent *SWS*** **5.2.1 Biomarker interpretation information** **5.2.1.1 Biomarker name** Vibration-controlled transient Elastography- (VCTE-) based Young's modulus (*E*) and equivalent shear wave speed (*SWS*) **5.2.1.2 Biomarker type** Diagnostic enrichment biomarker, adapted from the BEST Glossary **5.2.1.3 Analytical methods** Like SWE-based *SWS*, VCTE-based *E*, and equivalent *SWS,* is a non-invasive biomarker of hepatic fibrosis in NAFLD and is measured by a commercial device (Fibroscan, Echosens). In VCTE, an external probe is placed against the skin. The probe generates a mechanical push to generate shear waves in liver tissue. By following the A-mode line and TM image position, operators select a homogeneous liver area (**Figure 5.3**). The same external probe detects and processes the induced shear wave, and the machine displays *E* by default (and equivalent *SWS* when configured for US) results on the screen. Two probes are available for adults. The M probe is for subject with a probe-to-liver-capsule distance (PCD) below 30 mm while the XL probe is designed for patients with a PCD ≥ 30 mm. Recent VCTE software (SmartExam) automatically adjusts the region of measurement to the PCD value and recommends the appropriate probe to use. Measurements are not recommended when the PCD exceeds 45 mm. A third probe (S probe) is available for small patients. **5.2.1.4 Measurement units and limit(s) of detection** **Shear wave speed** (*SWS*) values are expressed in units of meters per second (m/s), and **Young's modulus** (*E*) values are expressed in units of kilopascals (kPa) (Ozturk et al. 2018a). The relationship between *SWS* and *E* is $E=3\rho(SWS)^2$, where ρ is equal to 1000 $kg \cdot m^{-3}$. Both values can be displayed; however, the default displayed value is Young's modulus, *E*. **Limits of detection** for VCTE-based SWS and *E* can range from 0.7 to 5.0 m/s, and from ~ 1.5 to 75.0 kPa, respectively (Fibroscan device user guide). **5.2.1.5 Biomarker interpretation and utility** Liver VCTE-based *E*, and equivalent *SWS*, is interpreted and used directly, without additional processing or modification. Based on a study showing that VCTE-based *E* values correlate with hepatic histologic fibrosis stage reference values (Siddiqui et al. 2019), the cutoff *E* values shown in **Table 5.3** have been proposed to classify histologic fibrosis stage and illustrate the covariation

between the biomarker and fibrosis stage. The actual cutoff values to be validated may differ from those listed in **Table 5.3**.

Notice that the AUROC values in **Table 5.3** are close to the expected values for a perfect biomarker under typical study conditions as explained in **Section 4.1**.

Table 5.3. VCTE-based cutoff values (Young's modulus) by histologic fibrosis stage

Histologic classification	AUROC	Cutoff value (kPa)	Sensitivity (%)	Specificity (%)	Negative likelihood ratio	Positive likelihood ratio
Any fibrosis (≥ Stage 1)	0.76 (0.64- 0.89)	8.6	0.53	0.87	0.37	0.93
Significant fibrosis (≥ Stage 2)	0.79 (0.74-0.83)	8.6	0.66	0.80	0.70	0.78
Advanced fibrosis (≥ Stage 3)	0.83 (0.79-0.87)	8.6	0.80	0.74	0.89	0.59
Cirrhosis (Stage 4)	0.93 (0.90-0.97)	13.1	0.89	0.86	0.99	0.39

Data in this table from Siddiqui et al, ***2019***

A common **actionable conclusion** (use case) for VCTE-derived *E*, and equivalent *SWS*, as a diagnostic enrichment biomarker in drug development trials is to enrich a population sample at screening for histologic fibrosis stage above a certain histologic fibrosis stage. For example, a VCTE liver stiffness cutoff at or near 8.6 kPa could be used to enrich a trial intending to enroll subjects with histologic fibrosis stages ≥ 2. It is also possible that a higher cutoff could be used to exclude subjects likely to have cirrhosis; an appropriate cutoff for this purpose has not yet been identified, and will be assessed in NIMBLE studies.

5.2.2 Analytical considerations

5.2.2.1 Method of measurement

VCTE-based *E*, and equivalent *SWS*, is acquired and calculated as described in **Section 5.2.1.3** (Sigrist et al. 2017).

5.2.2.2 Quality assurance and quality control

QA and QC procedures are described in several papers (Armstrong et al. 2013; Pang et al. 2014). QA/QC components suggested by the manufacturer include: (1) patients should be fasting for at least three hours before conducting FibroScan procedure; (2) Obtain at least 10 valid measurements; (3) if median E value is >= 7.1 kPa then IQR/median should be =< 30% (Boursier et. al. 2013).

5.2.2.3 Validation studies

The **measurement tool** is the standard VCTE-based *E,* and equivalent *SWS*, exam (**Section 5.2.1.3**).

The manufacturer of VCTE, Echosens, does not recommend or require device calibration. Thus, validation studies to date have not required device calibration. The manufacturer of VCTE, Echosens, does recommend periodic calibration of the vibration mechanism of VCTE probe once a year.

Studies have been conducted to assess **diagnostic accuracy** (Eddowes et al. 2019, Siddiqui et al. 2019) and **precision**, including same-operator test-retest repeatability and different-operator reproducibility (Vuppalanchi et al. 2018, Felicani et al. 2018; Leong et al. 2020; Serra et al. 2020).

Also, as part of the NIH subaward HHSN268201500021C, same-operator test-retest repeatability and different-operator reproducibility under QIBA profile suggestions, were reported to the RSNA QIBA leadership. The manuscript is under production. The unpublished results are provided in **Appendices 1** and **2**. The results include imaging techniques which use SWE and VCTE technology.

5.2.2.4 Final biomarker validation

NIMBLE studies will provide more comprehensive assessments of *SWS*-based liver stiffness precision and, using histology as the reference standard, accuracy.

	5.2.2.5 Additional considerations **Hepatic VCTE-based *E,* and equivalent *SWS,* acquisition and analysis has been standardized** in published validation studies (Siddiqui et al. 2019) and drug development trials (**Appendix 4**). **5.2.3 Supporting information** **5.2.3.1 Underlying biological process** VCTE-based *E,* and equivalent *SWS,* is a quantitative imaging biomarker of hepatic fibrosis (Sigrist et al. 2017), and hepatic fibrosis is one of the **underlying biological processes** contributing to the development and progression of NASH (Ozturk et al. 2020). **5.2.3.2 Clinical data to support the biomarker in its COU** VCTE-based liver stiffness has been used in numerous NASH drug development trials as a diagnostic enrichment biomarker, summarized in **Appendix 4**, from ***clinicaltrials.gov***. **5.2.3.3 Planned studies to support the biomarker and COU** The NIMBLE studies will address the gaps in knowledge for this COU described in **Section 7.5**. **5.2.3.4 Alternative comparator, current standard(s), or approaches** **Alternate comparator imaging biomarkers** to assess hepatic fibrosis include MRI-based biomarkers of hepatic fibrosis and inflammation, which are being assessed in NIMBLE studies (**Section 6**).
VCTE CAP	**5.3 VCTE-based CAP** **5.3.1 Biomarker interpretation information** **5.3.1.1 Biomarker name** Controlled Attenuation Parameter (CAP) **5.3.1.2 Biomarker type** Diagnostic enrichment biomarker, adapted from the BEST Glossary **5.3.1.3 Analytical methods** CAP is a non-invasive biomarker of hepatic steatosis in NAFLD. CAP is measured by the same device and probe used for VCTE-based *E,* and equivalent *SWS*. The attenuation of the ultrasound beam used during the examination is used to generate a CAP value using a proprietary algorithm. As with VCTE-based *E,* and equivalent *SWS*, S, M or XL probes can be used. **5.3.1.4 Measurement units and limit(s) of detection** CAP results are provided in dB/m unit. The technology can provide values between 100 to 400 dB/m *(Fibroscan User Guide*). **5.3.1.5 Biomarker interpretation and utility** Liver CAP is interpreted and used directly, without additional processing or modification CAP values can be useful to differentiate patients with from those without steatosis and also to diagnose patients with advanced steatosis (Ozturk et al. 2018b). Using biopsy-based steatosis grade as reference standard, several CAP cut-off points have been proposed to detect different histologic steatosis grades (**Table 5.4**). The **actionable conclusion** (use case) for CAP measurements as a diagnostic enrichment biomarker in drug development trials is to enrich clinical trial population samples by identifying histologic steatosis above a certain histologic grade. For example, a CAP cutoff at or near ~ 300 dB/m could be used to enrich a trial intending to enroll subjects with Brunt histologic steatosis grades ≥ S1 (Eddowes et al. 2019).

5.3.2 Analytical considerations

5.3.2.1 Method of measurement

CAP is acquired and calculated as described in **Section 5.3.1.3**.

5.3.2.2 Quality assurance and quality control

QA and QC procedures are described in the vendor instruction manual. A measurement gauge of 100% is recommended (meaning that at least valid 200 measurements have been obtained).

The operators can check the TM mode and A mode in the VCTE measurements (**Figure 5.3**).

Table 5.4 illustrates covariation between the biomarker and steatosis grade. The actual cutoff values to be validated may differ from those listed in **Table 5.4**.

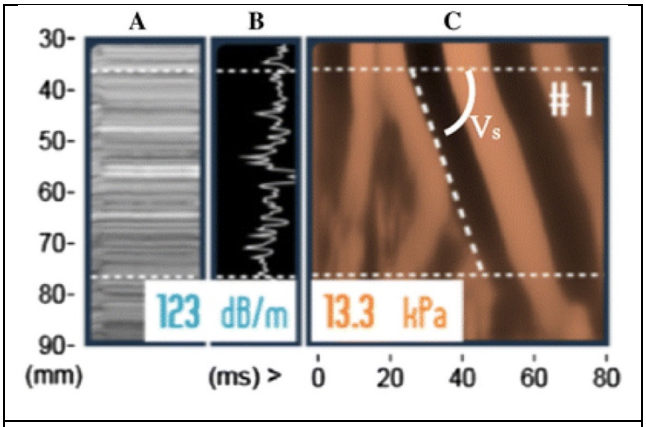


Figure 5.3. a) Time-motion (TM) mode **b)** Amplitude (A) mode. TM and A modes are used to locate ideal liver locations for measurement. **c)** Shear wave propagation image. y-axis is distance from skin, x-axis is time. Slope of the dashed line is shear wave speed (V_S). Tissue stiffness value is indicated in kPa. In the left panel, controlled attenuation parameter (CAP) value, which quantifies steatosis, is indicated in dB/m.

Table 5.4. Summary of VCTE-CAP-based cutoffs by steatosis grade

Study	N	Probe type	Histologic steatosis classification	Mean / Median CAP value (dB/m)	Optimal cut-off value (dB/m)	AUROC	AUROC LCI	AUROC UCI	Sensitivity (%)	Specificity (%)
(Friedrich-Rust et al. 2012)	46	M	≥ S1	241	N/A	N/A	N/A	N/A	N/A	N/A
			≥ S2	298	245	0.78	0.58	0.99	97	67
			≥ S3	319	301	0.72	0.57	0.86	76	68
(Myers et al. 2012)	153	M	≥ S1	299*	289	0.79	0.71	0.87	68	88
			≥ S2	319*	288	0.76	0.69	0.84	85	62
			≥ S3	320*	283	0.7	0.6	0.81	94	47
(Kumar et al. 2013)	63	N/A	≥ S1	213*	N/A	N/A	N/A	N/A	N/A	N/A
			≥ S2	284*	258	0.79	N/A	N/A	78.4	73.1
			≥ S3	291*	283	0.76	N/A	N/A	71.4	67.9
(Chan et al. 2014)	161	M	≥ S1	305*	263	0.97	N/A	N/A	91.8	93.7
			≥ S2	320*	263	0.86	N/A	N/A	96.9	67.7
			≥ S3	324*	281	0.75	N/A	N/A	100	53.1
(Karlas et al. 2014)	50	M	≥ S1	253	233.5	0.93	0.86	1	93	87
			≥ S2	321	268.5	0.94	0.88	0.99	97	81
			≥ S3	335	301.2	0.82	0.7	0.93	82	76
(Imajo et al. 2016)	142	M	≥ S1	262.9	236	0.88	0.8	0.95	82.3	91
			≥ S2	289.6	270	0.73	0.64	0.81	64.3	73.6
			≥ S3	304.9	302	0.7	0.58	0.83	64.3	73.6
(de Lédinghen et al. 2016)	261	M	≥ S1	264	N/A	N/A	N/A	N/A	N/A	N/A
			≥ S2	331	310	0.8	0.73	0.86	79	71
			≥ S3	336	311	0.66	0.59	0.72	87	47
(Lee et al. 2016)	183	N/A	≥ S1	265*	247	0.95	0.92	0.98	88.2	100
			≥ S2	313*	280	0.85	0.79	0.91	84.7	80
			≥ S3	322*	300	0.72	0.64	0.80	72.7	60.7
(Chan et al. 2017)	57	M	≥ S1	324*	266	0.94	0.86	0.98	91.1	87
			≥ S2	321*	266	0.80	0.69	0.88	91.1	87
			≥ S3	330*	267	0.69	0.57	0.79	100	40.6
		XL	≥ S1	339*	271	0.97	0.90	0.99	94.6	91.3
			≥ S2	345*	271	0.81	0.71	0.89	95.3	61.1
			≥ S3	345*	304	0.67	0.56	0.77	80	54.7
(Park et al. 2017)	104	M and XL	≥ S1	N/A	261	0.85	0.75	0.96	71.8	85.7
			≥ S2	N/A	305	0.70	0.58	0.82	63.3	68.8
			≥ S3	N/A	312	0.73	0.58	0.89	63.6	70.1
(Siddiqui et al. 2018)	393	M and XL	≥ S1	306*	285	0.76	0.64	0.89	80	77
			≥ S2	340*	311	0.70	0.64	0.75	77	57
			≥ S3	340*	306	0.58	0.51	0.64	80	40
(Garg et al. 2018)	76	XL	≥ S1	320*	323	0.75	0.61	0.89	58.6	83.3
			≥ S2	354*	336	0.74	0.62	0.86	73.9	75.5

			≥ S3	362*	357	0.82	0.73	0.91	100	77.8
(Chan et al. 2018)	101	M	≥ S1	293*	253	0.84	0.78	0.89	92.6	70.6
			≥ S2	327*	294	0.76	0.69	0.82	84.8	58.8
			≥ S3	330*	294	0.61	0.54	0.69	87.9	36.1
		XL	≥ S1	302*	279	0.91	0.85	0.94	82.8	88.2
			≥ S2	342*	303	0.78	0.71	0.84	79.5	64.7
			≥ S3	351*	325	0.65	0.58	0.72	75.8	54.4
(Runge et al. 2018)	55	M	≥ S1	N/A	260	0.77	0.63	0.87	90	60
			≥ S2	N/A	296	0.78	0.65	0.88	92.3	55.2
			≥ S3	N/A	334	0.78	0.65	0.88	77.8	76.1

5.3.2.3 Validation studies

The **measurement tool** is the standard CAP exam (**Section 5.3.1.3**).

The manufacturer of CAP, Echosens, does not recommend or require device calibration. Thus, validation studies to date have not required device calibration.

Validation studies have been conducted to assess diagnostic accuracy and measures of precision, including same-operator test-retest repeatability and different-operator reproducibility (Vuppalanchi et al. 2018)

5.3.2.4 Final biomarker validation

NIMBLE studies will provide more comprehensive assessments of liver CAP precision and, using histology as the reference standard, accuracy.

5.3.2.5 Additional considerations

Hepatic CAP acquisition and analysis has been standardized in published validation studies (Carvalhana et al. 2014; Eddowes et al. 2019), and in drug development trials (**Appendix 5**).

5.3.3 Supporting information

5.3.3.1 Underlying biological process

CAP is a quantitative imaging biomarker of hepatic steatosis (Sigrist et al. 2017), and hepatic steatosis is one of the **underlying biological processes** contributing to the development and progression of NASH (Ozturk et al. 2020).

5.3.3.2 Clinical data to support the biomarker in its COU

CAP has been used in numerous NASH drug development trials as a diagnostic enrichment biomarker, summarized in **Appendix 5**, from ***clinicaltrials.gov***.

5.3.3.3 Planned studies to support the biomarker and COU

The NIMBLE studies will address the gaps in knowledge for this COU described above, in **Section 7.5**.

5.3.3.4 Alternative comparator, current standard(s), or approaches

Alternate comparator imaging biomarkers to assess hepatic steatosis include MRI-based biomarkers of hepatic steatosis, which are being assessed in NIMBLE studies (**Section 6**).

6. Magnetic Resonance Imaging (MRI) Biomarker Information

This section contains biomarker interpretation information, analytical consideration, and supporting information for each of the following three categories of MRI-based candidate biomarkers:

- MRI-based proton density fat fraction (PDFF)
- MRE-based liver stiffness
- MRI-based corrected T1 imaging (cT1)

PDFF

6.1 PDFF

6.1.1 Biomarker interpretation information

6.1.1.1 Biomarker name

MRI-derived proton density fat fraction (MRI-PDFF)

6.1.1.2 Biomarker type

Diagnostic enrichment biomarker, adapted from the BEST Glossary

6.1.1.3 Analytical methods

MRI-PDFF is a non-invasive biomarker of hepatic steatosis in NAFLD. MRI-PDFF is calculated with magnitude-based or complex-based confounder-corrected chemical-shift-encoded MRI techniques.

In **complex-based MRI-PDFF imaging**, both real and imaginary images are used to automatically compute pixel-by-pixel parametric PDFF maps. Users draw ROIs on those maps. Mean PDFF values of those ROIs are reported.

In **magnitude-based MRI-PDFF imaging**, only magnitude images (which are automatically constructed from real and imaginary images when the images are acquired) are used to calculate PDFF values. This can be done either by users drawing ROIs on the magnitude images first, and then performing least-squares fitting of the ROI values to compute PDFF values for those ROIs, or by creating a pixel-by-pixel PDFF map first, and then having users draw ROIs on those maps. In both cases, mean PDFF values of those ROIs are reported.

Calculations for the complex- and for the magnitude-based MRI-PDFF methods both assume the same triglyceride model to account for cross-peak spectral interference across the multiple triglyceride resonances, and that of water (Hamilton et al. 2011). Each method is acquired with appropriate values of TR and flip angle to minimize T1 weighting, and both methods allow correction for T2* decay by acquiring data at multiple echo times. Additionally, MRI-PDFF acquired by the complex-based method is corrected for phase errors such as those caused by eddy currents and noise-related bias (Liu et al. 2007; Yu et al. 2011).

6.1.1.4 Measurement units and limit(s) of detection

MRI-PDFF is expressed in **units** of PDFF percentage (%). PDFF is calculated as the ratio of corrected triglyceride MRI signal, to the sum of corrected MRI signals from triglyceride and water.

Limits of detection for the MRI-PDFF methods are described in **Section 6.1.2.1**, and range from ~ 0 to 50% for magnitude-based MRI-PDFF, and from ~ 0 to 100% for complex-based MRI-PDFF.

6.1.1.5 Biomarker interpretation and utility

Raw calculated MRI-PDFF is interpreted and used directly, without additional processing or modification.

Estimates of MRI-PDFF acquired by complex- and magnitude-based methods agree strongly with

each other (Artz et al. 2015; Cunha et al. 2020; Haufe et al. 2017; Mamidipalli et al. 2020). Diagnostic accuracy has been reported to be high compared to MRS reference values, histologic steatosis grade reference values (Middleton et al. 2017; Middleton et al. 2018; Tang et al. 2013) , liver triglyceride concentration values, and phantom reference values.

Table 6.1. MRI-PDFF cutoff values by histologic steatosis grade

Histologic classification	AUROC	MRI-PDFF cutoff value (% PDFF)	Sensitivity (%)	Specificity (%)	Negative predictive value (%)	Positive predictive value (%)
Steatosis grade 0 vs. ≥ 1	0.99[a,b]	6.4[a,b]	97[a] 96[b]	100[a] 100[b]	71[a] 63[b]	100[a] 100[b]
Steatosis grade ≤ 1 vs. ≥ 2	0.83[a,b] 0.95[c] 0.87[d]	17.4[a,b] 16.3[c] 17.5[d]	61[a] 54[b] 83[c] 74[d]	90[a] 81[b] 90[c] 90[d]	61[a] 54[b] 73[c] 41[d]	90[a] 81[b] 95[c] 97[d]
Steatosis grade ≤ 2 vs. ≥ 3	0.89[a,b] 0.96[c] 0.79[d]	22.1[a,b] 21.7[c] 23.3[d]	68[a] 74[b] 84[c] 60[d]	91[a] 81[b] 90[c] 90[d]	90[a] 90[b] 94[c] 65[d]	72[a] 56[b] 76[c] 88[d]

a - Tang et al, *Radiology* **2013**; 267:422-431 (raw performance parameters)
b - Tang et al, *Radiology* **2013**; 267:422-431 (cross-validated performance parameters)
c - Middleton et al, *Gastroenterology* **2017**; 153:753-761 (cross-validated performance parameters)
d - Middleton et al, *Hepatology* **2018**; 67:858-872 (cross-validated performance parameters)

Based on studies showing that MRI-PDFF values correlate with hepatic histologic steatosis grade reference values (Middleton et al. 2017; Middleton et al. 2018; Tang et al. 2013) the cutoff MRI-PDFF values shown in **Table 6.1** have been proposed to classify histologic steatosis grade. Notice that the AUROC values in **Table 6.1** are close to the expected values for a perfect biomarker under typical study conditions as explained in **Section 4.1**.

A common **actionable conclusion** (use case) for using MRI-PDFF as a diagnostic enrichment biomarker in drug development trials is to enrich a population sample at screening for histologic steatosis grade at or above a certain histologic steatosis grade. For example, an MRI-PDFF cutoff at or near 6.4% could be used to enrich a trial intending to enroll subjects with histologic steatosis grade ≥ 1. Similarly, since steatosis is one of the three components of the NAS, MRI-PDFF could be used in combination with other biomarkers to identify patients more likely to have higher NAS.

6.1.2 Analytical considerations

6.1.2.1 Method of measurement

MRI-based PDFF is acquired and calculated as described in **Section 6.1.1.3.**

6.1.2.2 Quality assurance and quality control

QA and QC procedures are described in the Quantitative Imaging Biomarkers Alliance (QIBA) Profile for MRI-PDFF.

6.1.2.3 Validation studies

The **measurement tool** is the standard MRI-PDFF exam, either magnitude- or complex based (**Section 6.1.2.1**).

No explicit **device calibrations** are required for MRI-PDFF sequences apart from the standard general-purpose MRI calibrations required for all MRI scanners. Thus, validation studies to date have not required device calibration.

Validation studies have been conducted to assess diagnostic accuracy, and measures of precision, including inter- and intra-exam test-retest repeatability (Negrete et al. 2014) and different-manufacturer and different-field-strength reproducibility. Linearity, bias, and precision have been reported to be excellent supporting the following conclusion (Yokoo et al. 2018).

	Twenty-three studies (1679 participants) were selected for linearity and bias analyses and 11 studies (425 participants) were selected for precision analyses. MR imaging-PDFF was linear with MR spectroscopy-PDFF (R^2 = 0.96). Regression slope (0.97; P < .001) and mean Bland-Altman bias (20.13%; 95% limits of agreement: 23.95%, 3.40%) indicated minimal underestimation by using MR imaging-PDFF. MR imaging-PDFF was precise at the region-of-interest level, with repeatability and reproducibility coefficients of 2.99% and 4.12%, respectively. Field strength, imager manufacturer, and reconstruction method each had minimal effects on reproducibility. **6.1.2.4 Final biomarker validation** NIMBLE studies will provide more comprehensive assessments of MRI-based liver PDFF precision and, using histology as the reference standard, accuracy. **6.1.2.5 Additional considerations** **MRI-PDFF image acquisition and analysis has been standardized** in published MRI-PDFF validation studies, drug development trials (**Appendix 6**), and as per the QIBA MRI-PDFF Profile in development. Complex-based MRI-PDFF image acquisition and analysis have further been standardized as product sequences by MRI vendors for use on their platforms. **6.1.3 Supporting information** **6.1.3.1 Underlying biological process** MRI-PDFF is a quantitative imaging biomarker of hepatic steatosis, and hepatic steatosis is one of the **underlying biological processes** contributing to the development and progression of NASH. **6.1.3.2 Clinical data to support the biomarker in its COU** MRI-PDFF liver has been used in numerous NASH drug development trials as a diagnostic enrichment biomarker, summarized in **Appendix 6**, from ***clinicaltrials.gov***. **6.1.3.3 Planned studies to support the biomarker and COU** The NIMBLE studies will address the gaps in knowledge for this COU described above in **Section 7.5**. **6.1.3.4 Alternative comparator, current standard(s), or approaches** **Alternate comparator imaging biomarkers** to assess hepatic steatosis include ultrasound-based biomarkers of hepatic steatosis, which are being assessed in NIMBLE studies (**Section 5**), and magnetic resonance spectroscopy-PDFF (MRS-PDFF) which is not included in the NIMBLE studies because of difficulties in its widespread implementation in drug development trials.
MRE stiff-ness	**6.2 MRE-based magnitude of the complex shear modulus ('stiffness')** **6.2.1 Biomarker interpretation information** **6.2.1.1 Biomarker name** Magnetic resonance elastography- (MRE-) based magnitude of the complex shear modulus ('stiffness') *The magnitude of the complex shear modulus of the liver is commonly referred to as 'MRE-derived liver stiffness', or just as 'liver stiffness', and these terms will be used subsequently in this document.* **6.2.1.2 Biomarker type** Diagnostic enrichment biomarker, adapted from the BEST Glossary **6.2.1.3 Analytical methods** Liver stiffness is associated with histologic liver fibrosis stage and is a promising biomarker of liver fibrosis. Liver stiffness is calculated with **2D MRE or 3D MRE acquisition pulse sequences and data reconstruction methods**.

Both methods use an acoustic driver system with a vibrating passive driver placed on the subject's abdomen to generate continuous shear waves within the liver.

2D MRE acquisition pulse sequences and data reconstruction methods apply a 2D phase-contrast pulse sequence using a motion-encoding gradient in the through-slice direction to detect within-slice tissue motion caused by the shear waves. Postprocessing is performed using a 2D inversion algorithm to generate parametric map images (called elastograms) of liver stiffness as well as wave images and confidence maps.

3D MRE acquisition pulse sequences and data reconstruction methods use similar acquisition hardware and software as 2D MRE, except that 3D MRE acquisition sequences detect tissue motion within slices and across adjacent slices using motion-encoding gradients in three directions. Postprocessing is performed using a 3D inversion algorithm to generate elastograms as well as wave images and confidence maps.

For both 2D and 3D MRE, per-subject weighted values of the liver stiffness are calculated for acquired slices through the widest part of the liver that show good wave propagation.

6.2.1.4 Measurement units and limit(s) of detection

MRE-derived liver stiffness is expressed in **units** of kilopascals (kPa).

Limits of detection for MRE-derived liver stiffness range from ~ 1 to over 20 kPa.

6.2.1.5 Biomarker interpretation and utility

Raw calculated liver stiffness values are interpreted and used directly, without additional processing or modification.

Cutoff 2D MRE-derived liver stiffness values have been proposed to classify histologic fibrosis stage (**Table 6.2**) (LOI on MRE-derived liver stiffness submitted to FDA by Resoundant on 04 May **2020**, accepted 07 Aug **2020**, reference number is DDTBMQ000099).

Table 6.2. 2D MRE-derived cutoff values by histologic fibrosis stage from Resoundant LOI

Fibrosis stage	**AUROC**	**Cutoff value** (kPa)	**Sensitivity** (%)	**Specificity** (%)	**PPV** (%)	**NPV** (%)
≥ F2	0.92 (0.76-1)	3.3	86 (65-100)	90 (60-100)	89 (65-100)	86 (54-100)
≥ F3	0.95 (0.83-1)	3.9	87 (71-100)	91 (72-100)	83 (48-100)	88 (3-100)
F4	0.95 (0.85-1)	4.8	89 (73-100)	92 (81-100)	72 (27-97)	95 (72-100)

The AUROC values in **Table 6.2** are close to the expected values for a perfect biomarker under typical study conditions, as explained in **Section 4.1**.

A common **actionable conclusion** (use case) for 2D MRE-derived liver stiffness as a diagnostic enrichment biomarker in drug development trials is to enrich a population sample at screening for histologic fibrosis stage above, or below a certain histologic fibrosis stage (i.e., an MRE liver stiffness cutoff at or near 3.3 kPa could be used to enrich a trial intending to enroll subjects with histologic fibrosis stages ≥ 2). It is also possible that a higher cutoff could be used to exclude subjects likely to have cirrhosis; an appropriate cutoff for this purpose has not yet been identified and will be assessed in NIMBLE studies.

6.2.2 Analytical considerations

6.2.2.1 Method of measurement

2D and 3D MRE liver stiffness is acquired and calculated as described in **Section 6.2.1.3**.

6.2.2.2 Quality assurance and quality control

QA and QC procedures are described in the Quantitative Imaging Biomarkers Alliance (QIBA) Profile for liver MRE (MR Elastography Biomarker Committee. MR Elastography of the Liver, Quantitative Imaging Biomarkers Alliance. Profile Stage: Consensus. QIBA, June 6, **2019**. Available from: http://qibawiki.rsna.org/index.php/Profiles).

MRE-derived liver stiffness values depend on the spatial fidelity of the acquired phase images;

therefore, field of view and image linearity should be assessed and confirmed on an ongoing basis using MRI scanner-recommended procedures (MR Elastography Biomarker Committee. MR Elastography of the Liver, Quantitative Imaging Biomarkers Alliance. Profile Stage: Consensus. QIBA, June 6, **2019**. Available from: http://qibawiki.rsna.org/index.php/Profiles). Technical failures such as faulty synchronization of the driver system or incorrect driver frequency settings can cause incorrect measurement.

6.2.2.3 Validation studies

The **measurement tool** is the standard MRE exam (**Section 6.2.1.3**).

No explicit **device calibrations** are required for MRE sequences apart from the standard general-purpose MRI calibrations required for all MRI scanners. Thus, validation studies to date have not required device calibration.

Validation studies have been conducted for **2D MRE-derived liver stiffness** to assess diagnostic accuracy (Chen et al. 2017; Cui et al. 2015; Furlan et al. 2020; Loomba et al. 2014; Loomba et al. 2016; Morisaka et al. 2017; Morisaka et al. 2018); agreement across techniques of acquisition (Bae et al. 2018; Cunha et al. 2018; Forsgren et al. 2015; Morin et al. 2018; Trout et al. 2016; Venkatesh et al. 2014; Wagner et al. 2016; Wang et al. 2018; Yasar et al. 2016; Zhan et al. 2019), and measures of precision (Kim et al. 2020; Schwimmer et al. 2017; Wang et al. 2017). In a meta-analysis of diagnostic performance in 12 studies of 697 patients who underwent 2D MRE, Singh et al. 2015 reported areas under the receiver operating curve ranging from 0.84 to 0.92. A separate meta-analysis of test-retest repeatability of 2D MRE by Serai et al. 2017 supports the following conclusion for 2D MRE, which has been adopted in the QIBA Profile for liver MRE (MR Elastography Biomarker Committee. MR Elastography of the Liver, Quantitative Imaging Biomarkers Alliance. Profile Stage: Consensus. QIBA, June 6, **2019**. Available from: http://qibawiki.rsna.org/index.php/Profiles):

> *A measured change in hepatic stiffness of 19% or greater, at the same site and with use of the same equipment and acquisition sequence, is inferred to indicate that a true change in stiffness has occurred with 95% confidence*

Validation studies have also been conducted for **3D MRE-derived liver stiffness** to assess diagnostic accuracy for classification of histologic fibrosis stage (Allen et al. 2020; Loomba et al. 2016; Morisaka et al. 2017) and measures of precision including test-retest repeatability (Kim et al. 2020; Wang et al. 2017).

6.2.2.4 Final biomarker validation

NIMBLE studies will provide more comprehensive assessments of 2D and 3D MRE-stiffness precision and, using histology as the reference standard, accuracy.

6.2.2.5 Additional considerations

Standardized hepatic 2D and 3D MRE image acquisition and analysis is described in published studies (**Section 6.2.2.3**) and in drug development clinical trial documentation (**Appendix 7**), and for 2D MRE image acquisition in the QIBA MRE Profile (MR Elastography Biomarker Committee. MR Elastography of the Liver, Quantitative Imaging Biomarkers Alliance. Profile Stage: Consensus. QIBA, June 6, **2019**. Available from: http://qibawiki.rsna.org/index.php/Profiles).

6.2.3 Supporting information

6.2.3.1 Underlying biological process

MRE-derived liver stiffness is a quantitative imaging biomarker of hepatic fibrosis, and hepatic fibrosis is one of the **underlying biological processes** contributing to the development and progression of NASH.

	#### 6.2.3.2 Clinical data to support the biomarker in its COU MRE-derived liver stiffness has been used in numerous NASH drug development trials as a diagnostic enrichment biomarker, summarized in **Appendix 7**, from ***clinicaltrials.gov***. #### 6.2.3.3 Planned studies to support the biomarker and COU The NIMBLE studies will address the gaps in knowledge for this COU described in **Section 7.5**. #### 6.2.3.4 Alternative comparator, current standard(s), or approaches **Alternate comparator imaging biomarkers** to assess hepatic fibrosis include ultrasound-based biomarkers of hepatic fibrosis and inflammation, which are being assessed in NIMBLE studies (**Section 5**).
cT1	## 6.3 cT1 ### 6.3.1 Biomarker interpretation information #### 6.3.1.1 Biomarker name MRI-based iron corrected T1 relaxation time (cT1) #### 6.3.1.2 Biomarker type Diagnostic enrichment biomarker, adapted from the BEST Glossary #### 6.3.1.3 Analytical methods MRI-based cT1 is a non-invasive quantitative imaging biomarker of inflammation and fibrosis. It is measured by a product abdominal non-contrast breath-hold MRI pulse sequence and data reconstruction method developed by Perspectum Diagnostics Ltd. This method measures liver T1 relaxation values using a Modified Look Locker Inversion (MOLLI) recovery sequence, with or without cardiac gating, and liver T2* relaxation values using a multi-echo spoiled-gradient acquisition. The T1 values are thought to reflect liver water content, which is increased in inflamed and/or fibrotic liver. Thus, longer T1 values suggest the presence of inflammation and/or fibrosis. Since T1 values are affected by the presence of iron and by field strength, there is a need to correct raw acquired MRI-based T1 times for these confounding effects. To accomplish this, acquired raw T1 values (**Section 6.3.1.3**) are corrected using a published algorithm (Tunnicliffe et al. 2017) for the presence of iron, using the measured T2* values as a noninvasive indicator of iron content (Wood et al. 2005). The algorithm also corrects for field-strength effects to produce field-strength independent cT1 values. #### 6.3.1.4 Measurement units **MRI-based cT1** is expressed in **units** of milliseconds (ms). #### 6.3.1.5 Biomarker interpretation and utility MRI-based cT1 can be is interpreted and used directly, without additional processing or modification. This application focuses on the unconverted cT1 value. A cutoff of 800 ms for MRI-based cT1 has been reported to achieve 82% enrichment to distinguish potential at-risk clinical trial subjects with NAS<4 or F<2, from subjects NAS > 4 and F ≥ 2 (and also NAS < 4 or F < 2 or F = 4, from subjects with NAS ≥ 4 and F2-F3) in validation datasets (Perspectum LOI submitted 18 September 2018, Tables 2 and 3; response provided by FDA on 19 January 2019; DDTBMQ000051). Additionally, cT1 values in the liver have been reported to correlate with both inflammation and fibrosis histology (Banerjee et al. 2014). A common **actionable conclusion** (use case) for MRI-based cT1 as a diagnostic enrichment biomarker in drug development trials is to enrich a population sample at screening for histologic fibrosis and inflammation stages above, or below a certain values (i.e., a MRI-based cT1 cutoff at or near 800 ms could be used to enrich a trial intending to enroll subjects with NAS ≥ 4 and F ≥ 2).

6.3.2 Analytical considerations

6.3.2.1 Method of measurement

MRI-based cT1 is acquired and calculated as described in **Section 6.3.1.3.**

6.3.2.2 Quality assurance and quality control

QA procedures rely on product sequences that are available from MRI scanner manufacturers, and documentation and necessary training from Perspectum.

QC procedures for those product sequences are provided by MRI scanner manufacturers.

QC for the cT1 post-processing step is explained in documentation provided by Perspectum. In addition, Perspectum provides for automated QC checking with regard to acquisition parameters, shim box settings, and that data was acquired with a supported scanner, and provides error box messages for a variety of errors. They also provide guidance for real-time checking of results by analysts with regard to patient preparation and position, motion, shim artifact, field heterogeneity, differences across adjacent slices, low signal to noise ratio, presence of contrast, subject position in scanner, and fat suppression (*Perspectum LOI*).

6.3.2.3 Validation studies

The **measurement tools** are the product abdominal non-contrast breath-hold MRI pulse sequence and data reconstruction method developed by Perspectum.

No explicit **device calibrations** are required for these product sequences apart from the standard general-purpose MRI calibrations required for all MRI scanners. Thus, validation studies to date have not required device calibration.

Validation of the currently offered, FDA-approved cT1 package from Perspectum was based in part on separate training and validation cohorts (*Perspectum LOI*). The training cohort consisted of 102 adult subjects in the U.K. from two separate studies (53 subjects from the RIAL-NICOLA trial, and 49 subjects from the CALM trial). All subjects underwent cT1 imaging, and had histology from liver biopsy assessed. The 800 ms cutoff came from that data. The validation cohort came from 135 subjects enrolled in a study performed in the United States, all of whom also underwent cT1 imaging and liver biopsy. That study provided the data to support the 82% enrichment that was found for cT1 imaging.

In a study of 61 adults imaged on 5 different scanners, Bachtiar et al. 2019 report for cT1, for repeatability and cross-platform reproducibility, coefficients of variation of 1.7% and 3.3%, biases -7.5 and 6.5 ms, and 95% limits of agreement of (-53.6, 38.5 ms) and (-76.3, 89.2 ms), respectively.

As part of the LOI for cT1, Perspectum summarized results comparing cT1 values resulting from the K143020 and K172685 cT1 calculation products and bench testing, demonstration that results are platform-independent for two different MRI platforms, and that results are standardized to 3T.

6.3.2.4 Final biomarker validation

NIMBLE studies will provide more comprehensive assessments of cT1 precision and, using histology as the reference standard, accuracy.

6.3.2.5 Additional considerations

MRI-based cT1 image acquisition and analysis has been standardized in published validation studies and drug development trials.

6.3.3 Supporting information

6.3.3.1 Underlying biological process

MRI-based liver cT1 values have been reported to correlate with disease activity (inflammation and ballooning), and with fibrosis histology (Banerjee et al. 2014). cT1 is thought to characterize tissue by the proportion of the extracellular fluid present within the liver tissue (*Perspectum LOI*).

6.3.3.2 Clinical data to support the biomarker in its COU

MRI-based cT1 has been used in NAFLD/NASH drug development trials (Harrison et al. 2018; Harrison et al. 2020), but to our knowledge, not yet as a diagnostic enrichment biomarker.

6.3.3.3 Planned studies to support the biomarker and COU

The NIMBLE studies will address the gaps in knowledge for this COU described above in **Section 7.5**.

6.3.3.4 Alternative comparator, current standard(s), or approaches

Alternate comparator imaging biomarkers to assess hepatic fibrosis include ultrasound-based biomarkers of hepatic fibrosis, which are being assessed in NIMBLE studies (**Section 5**).

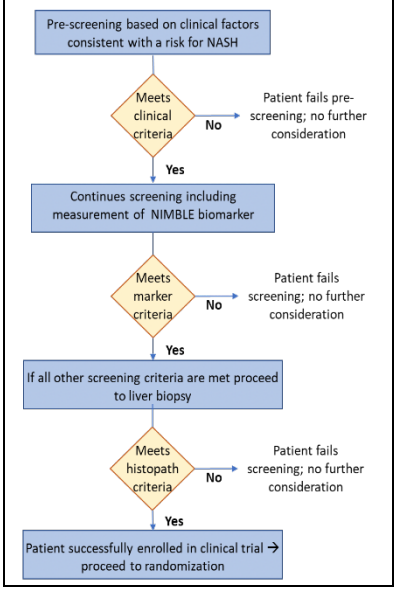
7. Clinical Considerations *(all candidate biomarkers)*

7.1 Use of biomarkers

The biomarkers described herein will be used as diagnostic enrichment biomarkers to help identify non-cirrhotic subjects at risk for NASH with liver fibrosis who will meet the criteria for enrollment in drug development trials, namely those subjects likely to have active NASH with stage 2 or 3 fibrosis. See **Appendices 3-7** for a list of trials on ***clinicaltrials.gov*** in which these biomarkers have been used for this purpose.

A **decision tree** is provided in **Figure 7.1** to explain how these biomarkers would be used in drug development trials.

Figure 7.1. Flow diagram for a diagnostic enrichment COU



7.2 Drug development setting

The biomarkers described herein may be used in drug development trials as diagnostic enrichment biomarkers as described above.

7.3 Clinical relevance

Using the biomarkers described herein as diagnostic enrichment biomarkers will reduce the frequency of screen-failures that involve the risk secondary to liver biopsy across Phase 2b-3 NASH drug development trials. The reduced frequency of screen failures will expedite drug development trials and accelerate the development of pharmacologic therapeutic interventions for NASH and reduce the number of unnecessary liver biopsies.

7.4 Benefits and Risks

The **benefits** of using the biomarkers described herein for drug development trials compared to biopsy are non-invasiveness, reduced study cost and risk, and potentially higher accuracy and precision.

The **risks** of using the biomarkers described herein for drug development trials vary, according to modality. For MRI-based biomarkers, risks include lack of point-of-care availability, lack of availability of any MRI in some regions, and lack of specialized software to acquire the required MRI exams at centers that do offer MRI otherwise. For ultrasound-based biomarkers, risks include the need for periodic calibration and technical service. Also, ultrasound-based exams are operator dependent, which might introduce operator variability.

7.5 Knowledge gaps, limitations, assumptions

Knowledge gaps being addressed in NIMBLE studies include:

1) MRI-based biomarker same-day and different-day repeatability for each of three MRI vendors (GE, Philips, Siemens) and both clinical field strengths (1.5T and 3T);
2) MRI-based biomarker reproducibility across vendors and field strengths;
3) Ultrasound-based biomarker same-day repeatability, and different-day different-operator reproducibility;
4) Cross-sectional accuracy of imaging-based biomarker values for histology classification; and
5) Longitudinal accuracy of change in imaging-based biomarker values for treatment response assessment.

8. Previous Qualification Interactions and Other Approvals *(all candidate biomarkers)*

8.1 Letters of Support (LOS) issued for this biomarker

No letters of support (LOS) have been submitted.

8.2 Discussion in a Critical Path Innovation Meeting (CPIM)

There have been no CPIM discussions.

8.3 Previous FDA Qualification given to this biomarker with DDT Tracking Record Number

a. 2D MRE is FDA cleared/approved (K083421, K121434, K140666, K183193, K201389).
b. Perspectum cT1 is FDA cleared/approved (K172685 through Perspectum, and K143020 through Mirada Medical)
c. Echosens Fibroscan is FDA cleared/approved (K123806, K150949, K150239, K160524, K172142, K173034, K181547, K200655, and K203273)

8.4 Qualification submissions to any other regulatory agencies with submission number

There have been no qualification submissions to any other regulatory agencies by Resoundant.

8.5 Prior or current regulatory submissions to Center for Biologics Evaluation and Research (CBER), Center for Drug Evaluation and Research (CDER), and Center for Devices and Radiological Health (CDRH). Provide 510(k)/PMA Numbers (https://www.fda.gov/drugs/biomarker-qualification-program/biomarker-qualification-submissions)

a. DDTBMQ000051 – Perspectum Letter of Intent to Qualify MRI-cT1; original submission date 09/16/2018
b. DDTBMQ000082 – Perspectum Letter of Intent to Qualify MRI-PDFF; original submission date 11/02/2018
c. DDTBMQ000084 – NIMBLE Letter of Intent to Qualify Circulating Biomarkers for Diagnosing Non-Alcoholic Steatohepatitis (NASH); original submission date 02/26/2019
d. DDTBMQ000099 – Resoundant Letter of Intent to Qualify 2D MRE; original submission date 04/13/2020

9. Attachments *(all candidate biomarkers)*

9.1 Publications

Please see **Appendix 8** for references.

9.2 Long-term goals

Development of the ultrasound- and MRI-based NIMBLE candidate biomarkers as diagnostic enrichment biomarkers for drug development trials in the NIMBLE studies by the Foundation for the National Institutes of Health (FNIH) to qualify quantitative imaging and circulating biomarkers for use in drug development trials.

9.3 Other supporting information

Appendix 1. Intra-operator and inter-operator agreement SWE-based SWS results from QIBA profile validation study

Appendix 2. Intra-operator and inter-operator agreement VCTE-based SWS results from QIBA profile validation study

Appendix 3. Clinical trials using SWE-based SWS as a diagnostic enrichment biomarker (*clinicaltrials.gov*)

Appendix 4. Clinical trials using VCTE-based SWS as a diagnostic enrichment biomarker (*clinicaltrials.gov*)

Appendix 5. Clinical trials using VCTE-based CAP as a diagnostic enrichment biomarker (*clinicaltrials.gov*)

Appendix 6. Clinical trials using MRI-PDFF as a diagnostic enrichment biomarker (*clinicaltrials.gov*)

Appendix 7. Clinical trials using MRE-based liver stiffness as a diagnostic enrichment biomarker (*clinicaltrials.gov*)

Appendix 8. References

10. Abbreviations *(all candidate biomarkers)*

3D	Three-dimensional
ARFI	Acoustic Radiation Force Impulse
AUROC	Area under receiver operating curve
CAP	Controlled attenuation parameter
CI	Confidence interval
COU	Context of use
cT1	Iron corrected T1 relaxation time
CV	Coefficient of variation
dB/m	Decibels per meter
E	Young's Modulus
FNIH	Foundation for the National Institutes of Health (NIH)
ICC	Intraclass correlation coefficient
IQR	Interquartile range
kPa	Kilopascals
LOI	Letter of intent
MMDI	Multi-modal direct inversion
MOLLI	Modified Look Locker Inversion
MRE	Magnetic resonance elastography
MRI	Magnetic resonance imaging
MRS	Magnetic resonance spectroscopy
NAS	NAFLD Activity Score
NIH	National Institutes of Health
NIMBLE	Non-Invasive Biomarkers of MetaBolic Liver DiseasE
NPV	Negative predictive value
PCD	Probe-to-liver-capsule distance

PDFF	Proton density fat fraction
PPV	Positive predictive value
QA	Quality assurance
QC	Quality control
QIBA	Quantitative Imaging Biomarkers Alliance
RC	Repeatability coefficient
ROI	Region of interest
SWE	Shear wave elastography
SWS	Shear wave speed
TE	Time to echo
TR	Repetition time
VCTE	Vibration-controlled transient elastography