Progress in development of a reasonably likely surrogate endpoint for MASH

Arun J. Sanyal M.D.

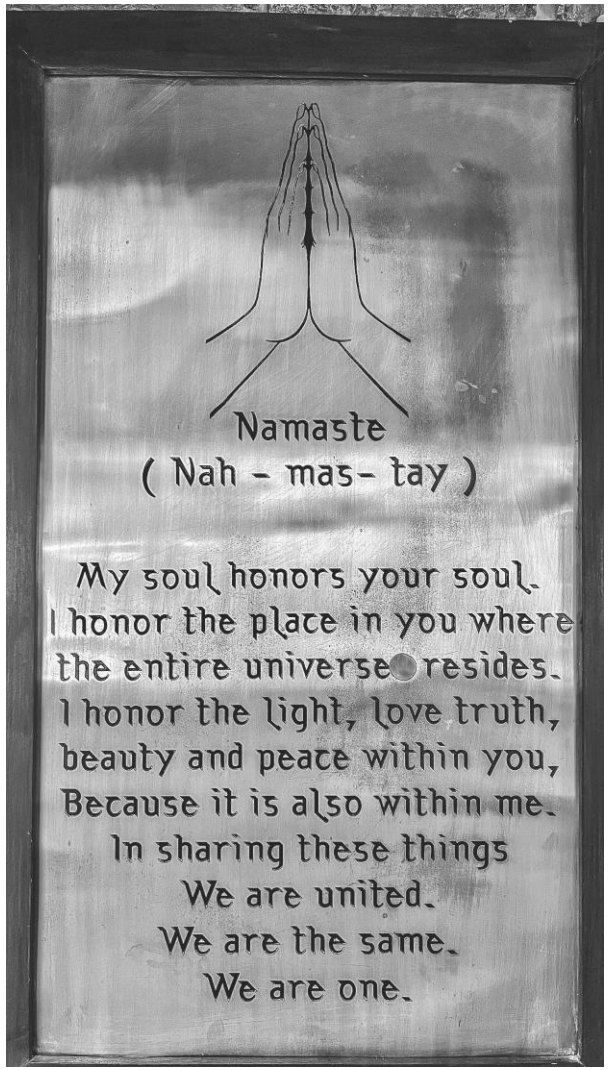
Z Reno Vlahcevic Professor of Medicine

Director, Stravitz-Sanyal Institute for Liver Disease and Metabolic Health

Virginia Commonwealth University School of Medicine

VCU

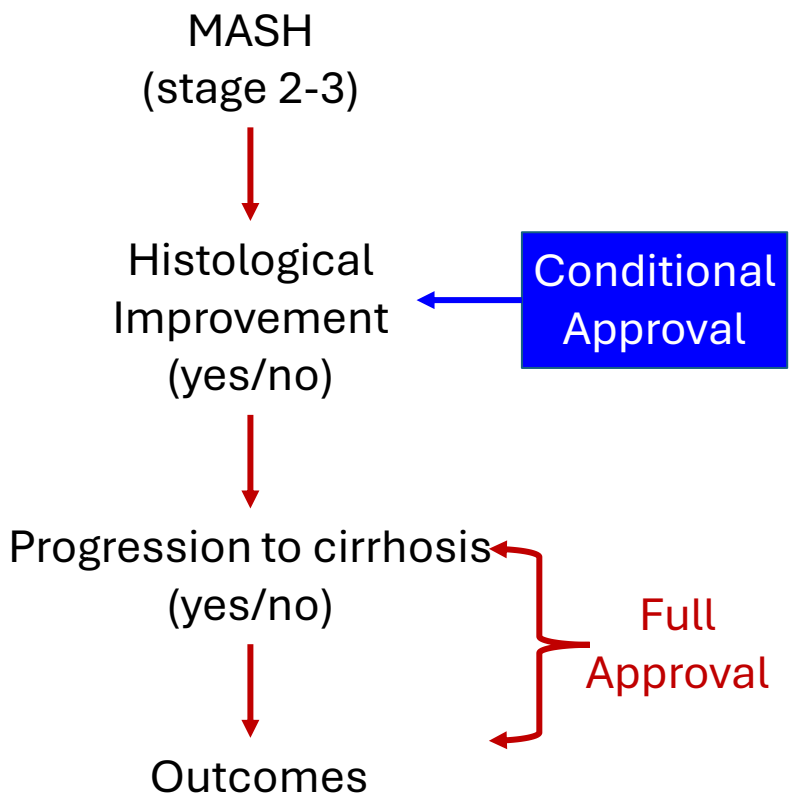
Stravitz-Sanyal Institute for
Liver Disease and Metabolic Health



I disclose the following financial relationship(s) with a commercial interest:

- Ownership interests: Durect, Tiziana, Genfit, Exhalenz, Northsea, Rivus, Inversago
- Consultant: Gilead, Intercept, Novartis, Novo Nordisk, Inventiva, Merck, Pfizer, Boehringer Ingelhiem, Bristol Myers Squibb, Eli Lilly, Genentech, Amgen, Alnylam, Regeneron, Thera Technologies, Madrigal, Salix, Malinckrodt, Gatehouse, Rivus, Siemens, Lipocine, 89 Bio, Astra Zeneca, Akero, Foresite, Mitopower, Histoindex, Path AI, Takeda
- Grant support to school: Gilead, Intercept, Novartis, Novo Nordisk, Inventiva, Eli Lilly, Genentech, Boehringer Ingelhiem, Bristol Myers Squibb

Current regulatory pathway- the status quo



- Conditional approval requires verification of clinical benefit
- Clinical benefit is defined as how a patient:
  - Feels
  - Functions
  - Survives
- The confirmatory trial may be a new study or continuation of phase 3 trial
- The phase 4 confirmatory trial is expected to be under way at time of accelerated “conditional” approval

Limitations of current paradigm

Technical
- Invasive
- Risk of serious morbidity/Mortality
- Pain

+

Analysis
- Biopsy quality
- Staining variability
- Sampling variability
- Intra-observer variability
- Inter-observer variability
- Drift over time

Since Bx is rarely used in clinical practice, translation from histological analytics to practice is difficult!

Another problem- Even after 5+ years, progression to cirrhosis is the only liver outcome seen in placebo/OCA arms in Regenerate

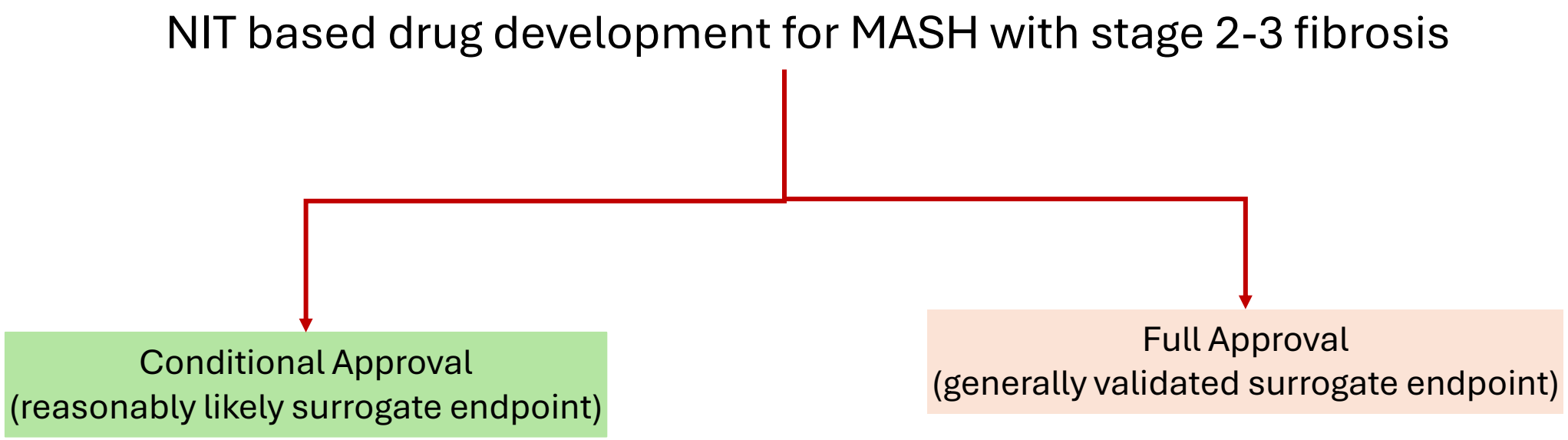
	Placebo (n=728)	OCA 10 mg (n=729)	OCA 25 mg (n=730)
Patients with any event, n	137	115	107
Rate of patients with any event, % (95% CI)	18.8 (16.0–21.9)	15.8 (13.2–18.6)	14.7 (12.2–17.4)
Mean rate difference (95% CI)		–3.0 (–6.9 to 0.8)	–4.2 (–8.0 to –0.4)
Stratified log-rank test P value vs placebo*		0.103	0.044
Stratified hazard ratio vs placebo (95% CI)†		0.814 (0.635–1.043)	0.772 (0.600–0.994)
Death	9 (1.2)	7 (1.0)	11 (1.5)
Liver transplant	0	1 (0.1)	1 (0.1)
MELD score ≥15	2 (0.3)	4 (0.5)	2 (0.3)
Hospitalization for complications of hepatic decompensation	0	1 (0.1)	2 (0.3)
Ascites secondary to cirrhosis requiring medical intervention	0	3 (0.4)	4 (0.5)
Progression to cirrhosis	126 (17.3)	99 (13.6)	87 (11.9)

*Statistical significance for clinical outcomes is set at $P<0.03$ (remained alpha from month 18 interim analysis). The log-rank test compared each treatment group to placebo stratified by baseline diabetes status and use of thiazolidinediones or vitamin E at baseline.
†The hazard ratio (95% CI) was calculated using a Cox regression model with baseline diabetes status and use of thiazolidinediones or vitamin E at baseline included in the model.
Abbreviations: MELD, model of end-stage liver disease; OCA, obeticholic acid.

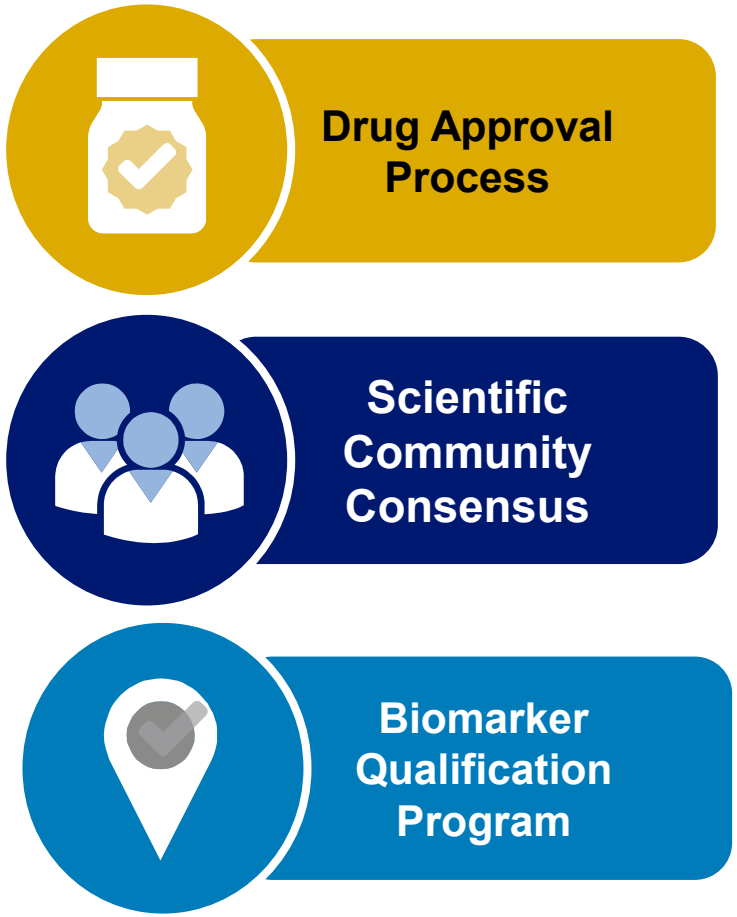
Sanyal et al, DDW 2024, MS in development

The technical and interpretive limitations of conventional histology based drug development and challenges in translation in to clinical management guidance provide a strong rationale to explore alternate approaches for MASH drug development!

Scope of this presentation



Multiple pathways for biomarker development from regulatory perspective



- Data Driven
- Subject to regulatory scrutiny
- More than one process can go on

Facilitating Biomarker Development: Strategies for Scientific Communication, Pathway Prioritization, Data-Sharing, and Stakeholder Collaboration; Published June 2016, Duke-Margolis Center for Health Policy

Liver Stiffness Measurement- LOI accepted for qualification- Aug 27, 2025

What is a reasonably likely surrogate endpoint (RLSE)?

An endpoint supported by strong mechanistic and/or epidemiologic rationale such that an effect on the *surrogate endpoint is expected to be correlated* with an **endpoint intended to assess clinical benefit in clinical trials**, but without sufficient clinical data to show that it is a validated surrogate endpoint.

- MECHANISTIC RATIONALE
- RELATIONSHIP OF NIT TO CLINICALLY MEANINGFUL OUTOCMES
- RELATIONSHIP OF CHANGE IN NIT TO CHANGE IN RISK OF OUTCOME
- LINKING THERAPEUTIC RESPONSE OF NIT TO CHANGE IN RISK OF CLINICALLY MEANINGFUL OUTCOMES

New Guidance recently accepted for Publication- Anania et al, Hepatology, DOI 10:1097, Aug 26, 2025

An academic perspective of how to measure clinical benefit to establish RLSE

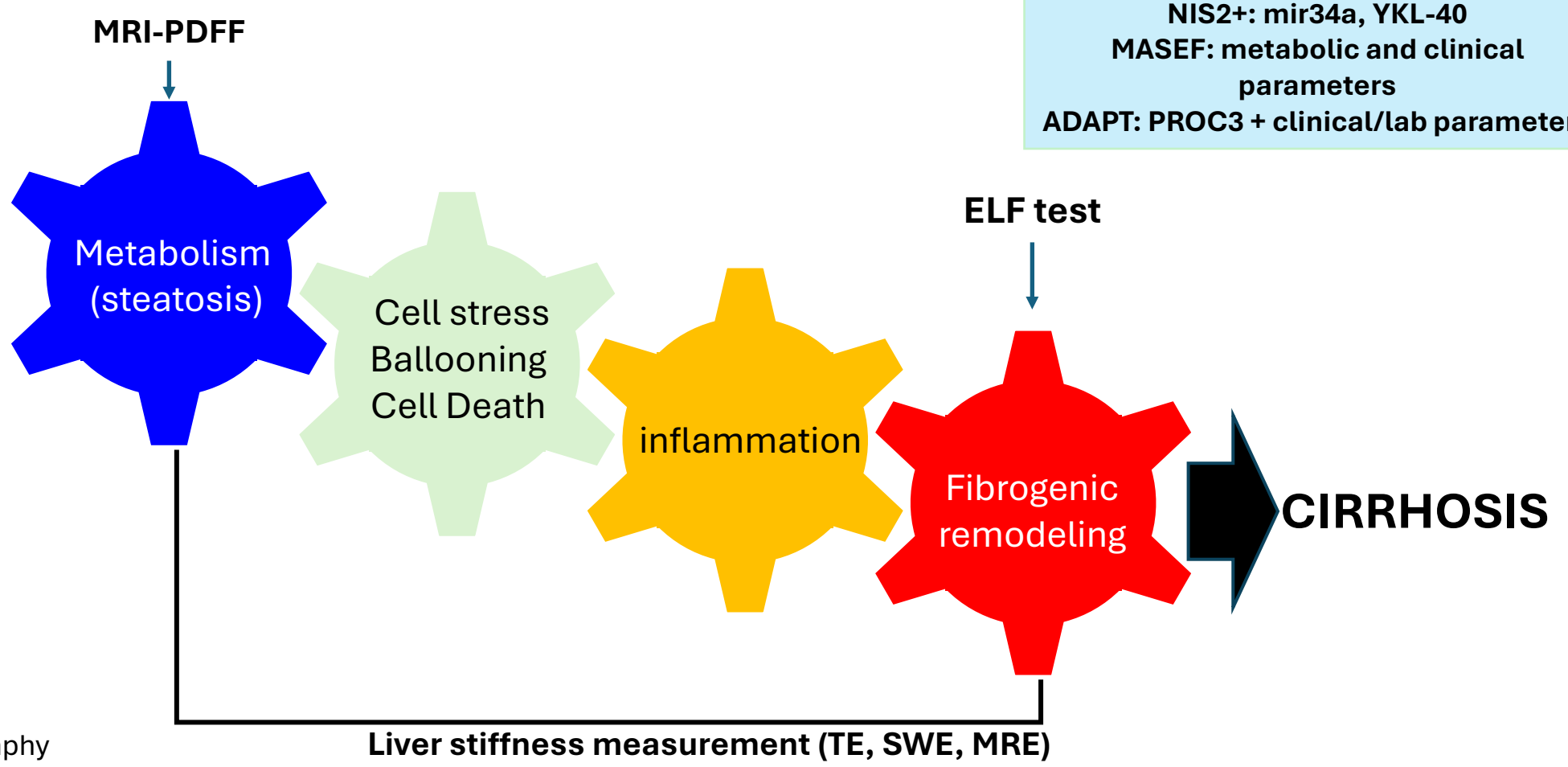
- Biologically relevant
- Easy to implement and use in clinical setting
- Safe
- Provide meaningful risk stratification
- Sensitive to change with good dynamic range
- Changes linked to change in disease status
- **Translatable to clinical trial implementation and clinical practice guidance**

Biological plausibility

Is the NIT measuring something in the disease pathway?

We already have tools that measure key elements in disease biology

Emerging NITs for NASH with fibrosis
NIS2+: mir34a, YKL-40
MASEF: metabolic and clinical parameters
ADAPT: PROC3 + clinical/lab parameters

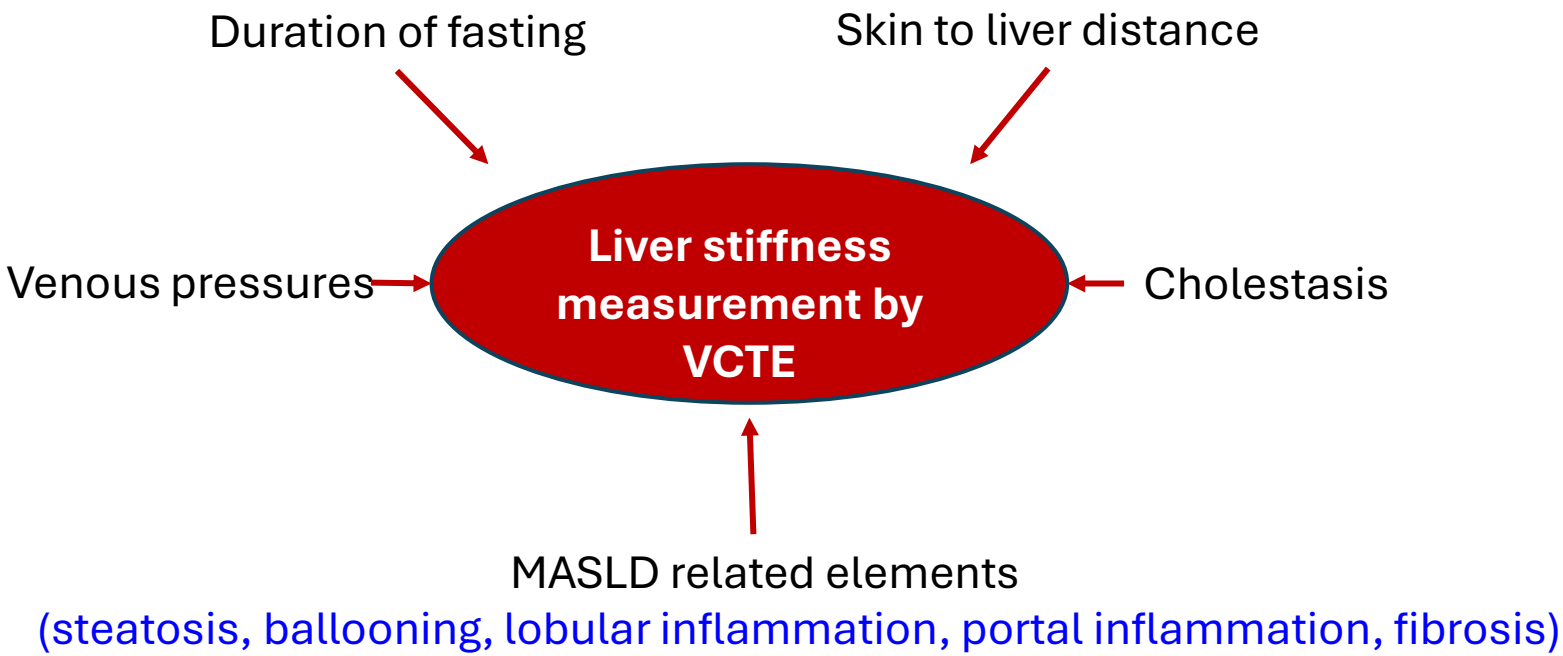


TE: transient elastography
SWE: shear wave elastography
MRE: magnetic resonance elastography
cT1: corrected T1

Adapted from Sanyal AJ. Nature Gastro Hepatol, 2019
Sanyal et al. Nature Medicine, 2023

Asking the tough questions about mechanistic basis of NIT change data!

- What are the elements that inform the NIT value?
- Which ones are linked to disease being studied?
- When the NIT changes, how much of the change is driven by disease related elements?



Vuppalanchi and Sanyal; Clin Gastro Hep, . 2015 Apr;13(4):780-2.

VCTE- how much of a change is a change?

	# subjects	# obs	Mean of Median SWS (m/s)	$RDC_{diff\text{-}day,\ diff\text{-}oper}$	Upper 95% confidence bound	$RC_{same\text{-}day,\ same\text{-}oper}$	Upper 95% confidence bound
Fibroscan/VCTE	39	39	1.641	35.6%	43.9%	19.6%	24.1%

Key Takeaway: Changes in shear wave speed as evaluated by VCTE **>35.6%** can be considered true change (with 95% confidence). However, the practical impact of this is "do we need 95% confidence for decision making"..usual threshold often lies in 80% range..e.g. LDL-C change or improvement in A1c vs outcomes

Pierce et al, 2024 Sep;312(3):e233094. doi: 10.1148/radiol.233094.

Precision Metrics

Sensitive to outliers- distribution of data are important to make clinical decisions

Primary metrics (as measured by primary analysis)

- Pooled diff-day, diff-operator reproducibility coefficient (pooled $RDC_{\text{diff-day, diff-operator}}$)

NHANES III

Analyte	Coefficient of variation
Bilirubin	23.4%
ALT	20.4%
AST	13.9%
GGT	13.8%
ALP	6.7%

N=1864, mean test 17.5 days apart

Lazo et al, Ann Intern Med. 2009 Apr 7;150(7):504. PMID: 18316753

LSM reflects biologically relevant elements of disease biology

N= 292

Fibrosis – stage 0-4	1.7 (1.3)
Stage 0 - %	24%
Stage 1 - %	24%
Stage 2 - %	20%
Stage 3 - %	23%
Stage 4 - %	9%
Definite NASH - %	58%

Mean time from Bx to Scan= 34 days

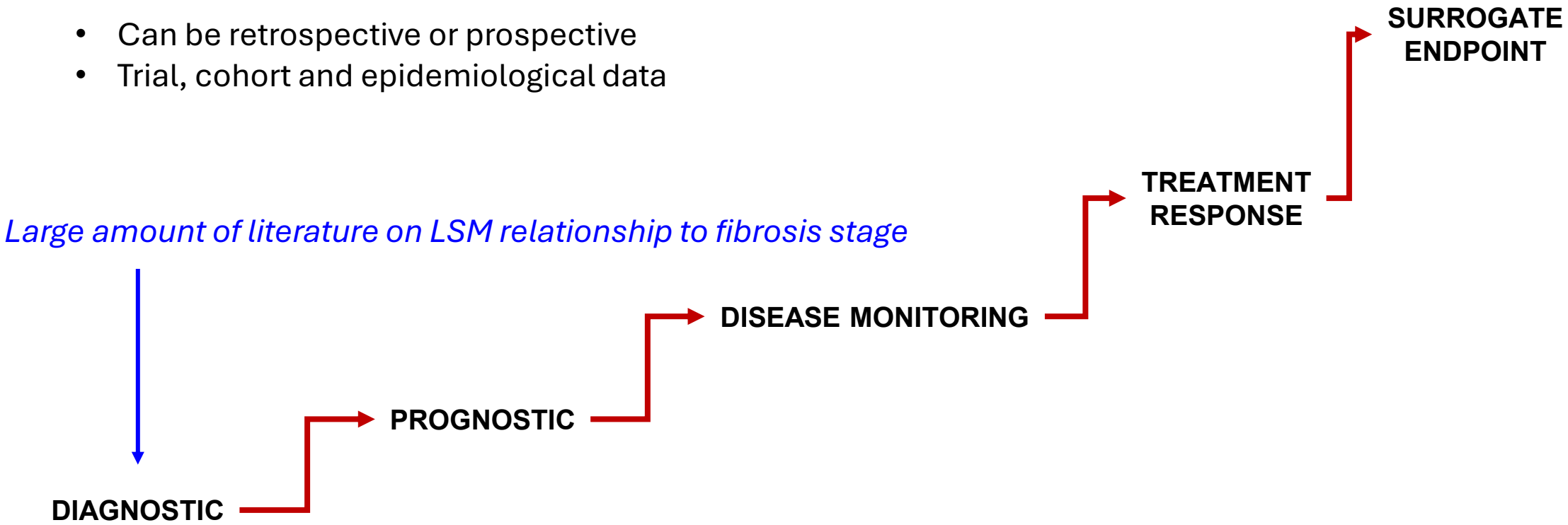
Outcome	Model	Covariate	Beta*	CI	P
Liver stiffness (kPa)	Model 1	Fibrosis stage	4.4	3.2, 5.6	<0.001
		Inflammation score	-0.4	-2.5, 1.7	0.70
	Model 2†	Fibrosis stage	5.1	3.7, 6.4	<0.001
		Steatosis score	-2.9	-4.4, -1.4	<0.001
		Inflammation score	1.9	-0.4, 4.2	0.10
		Ballooning score	-2.3	-4.5, -0.1	0.04
		Portal inflammation score	-1.7	-4.1, 0.7	0.15
CAP (dB/m)	Model 3†	Steatosis score	20	14, 27	<0.001

Siddiqui et al, CGH, 2019; 17:156-163

Linkage of NIT to clinical outcomes

FDA 2025- criterion # 2- what is the data to support that a NIT or group of NITs are adequate surrogates?

- Can be retrospective or prospective
- Trial, cohort and epidemiological data



LSM provides prognostic value

Liver Stiffness Thresholds to Predict Disease Progression and Clinical Outcomes in Advanced Fibrosis

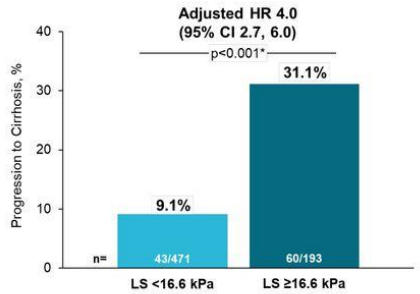
Aim:
To establish thresholds for Liver stiffness (LS) by vibration controlled transient elastography (VCTE) that predict progression to cirrhosis among patients with bridging fibrosis and hepatic decompensation among patients with cirrhosis due to NASH

Methods:
Prospective data from four randomized placebo-controlled trials of selonsertib (STELLAR-3; STELLAR-4) and simtuzumab (GS-US-321-0105; GS-US-321-0106) in participants with bridging fibrosis (n=664) and cirrhosis (n=734)

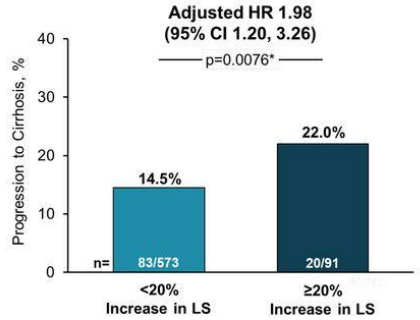
Conclusions:
The LS thresholds identified in this *prospective* study may be useful for risk stratification of patients with NASH in clinical trials and in clinical practice

Loomba et al. *GUT* 2022

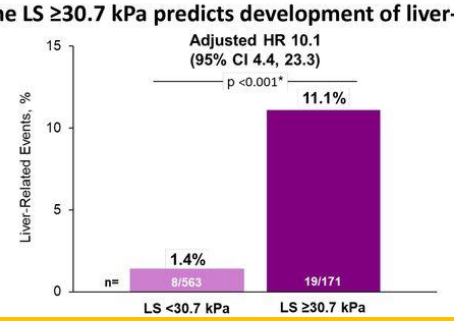
Baseline LS ≥16.6 kPa predicts progression to cirrhosis



5 kPa (and 20%) increase in LS predicts progression to cirrhosis

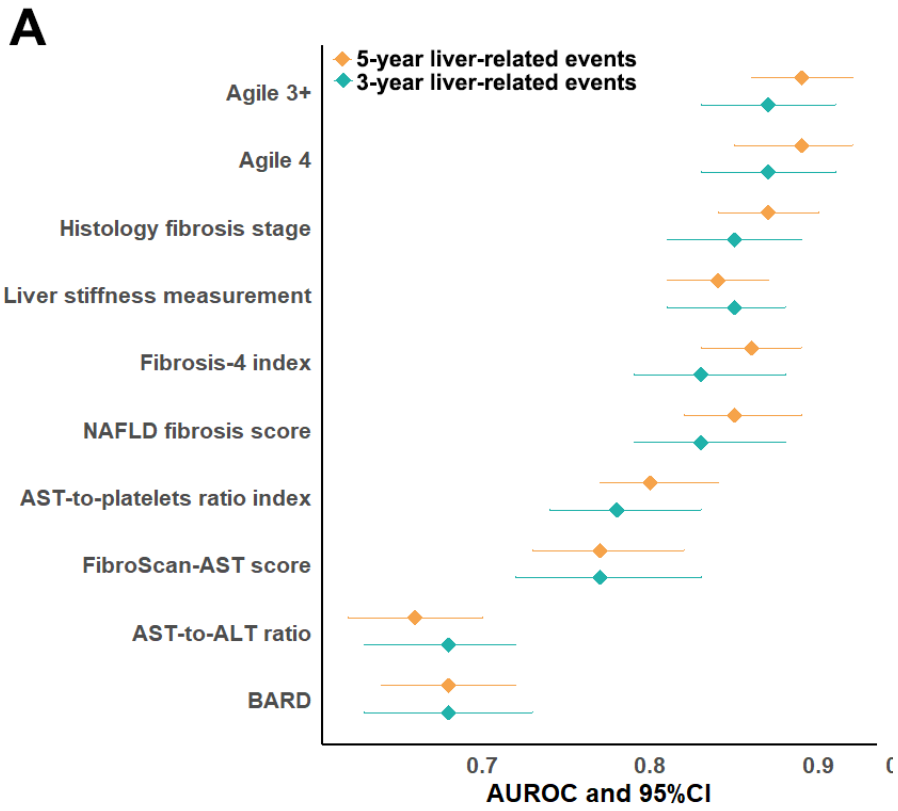


Baseline LS ≥30.7 kPa predicts development of liver-related events

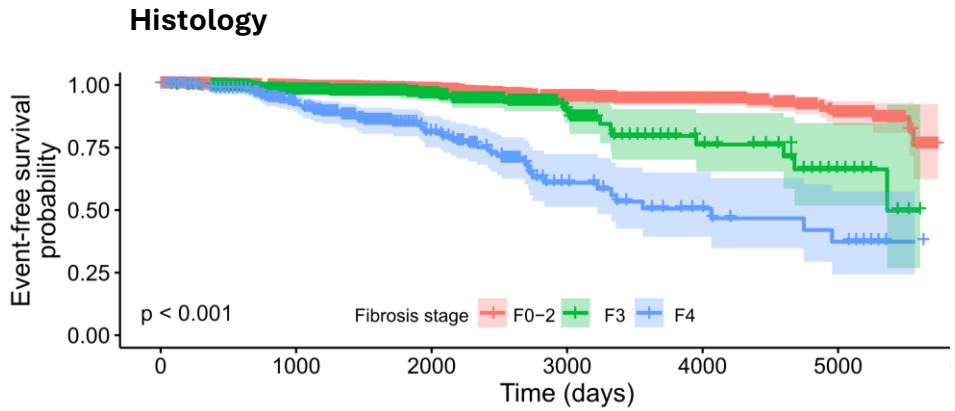


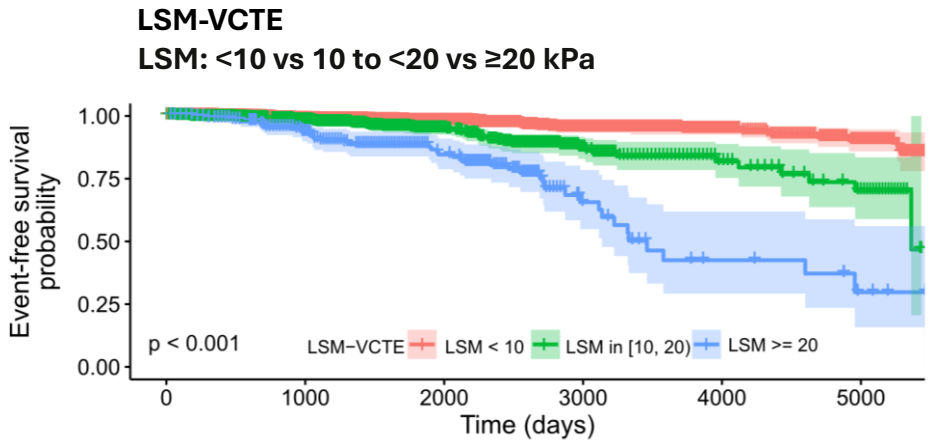
Loomba et al. Gut 2023;72:581-589

LSM predicts outcome risk



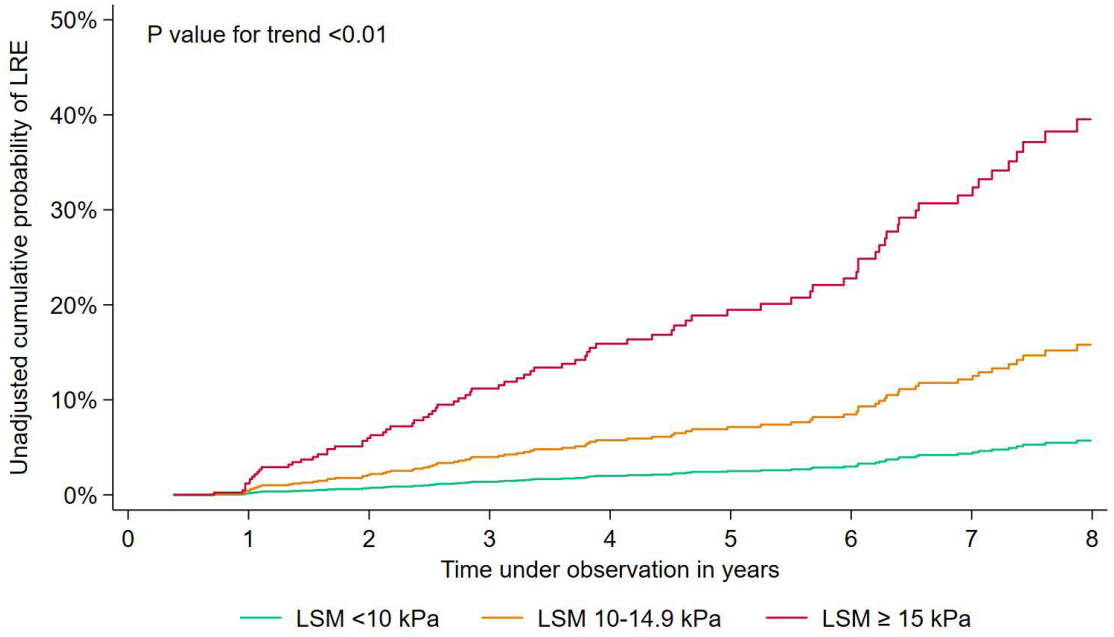
Lim et al, JAMA, 2024

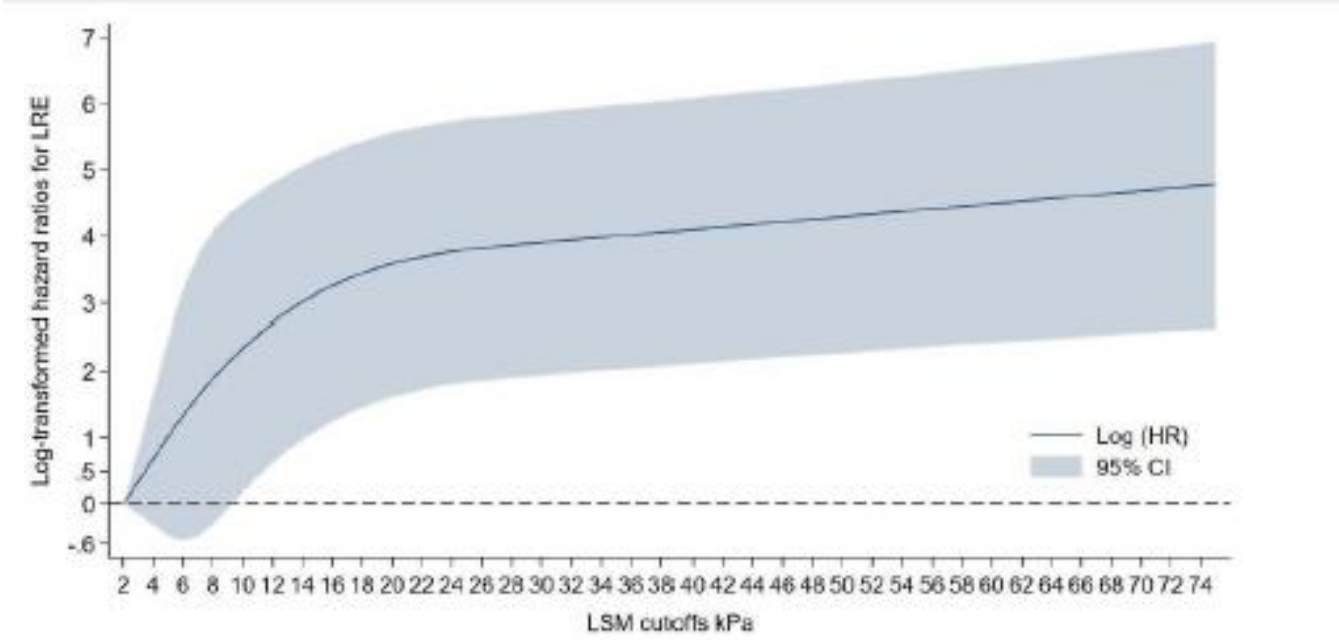




[1] Mozes et al.. Lancet Gastroenterol Hepatol 2023;8:704-713.

Prognostic utility of LSM- NASH CRN data





Which Patients- Aligning cutoffs

Comparison of liver-related clinical outcome risks for VCTE and FIB-4 cutoffs versus established ELF thresholds (Data from 6 placebo-controlled trials of MASH n= 2710)					
ELF		VCTE		FIB-4	
Established Cutoff	Clinical Outcome Risk (95% CI)	Comparable Cutoff (kPa)	Clinical Outcome Risk (95% CI)	Comparable Cutoff	Clinical Outcome Risk (95% CI)
Established ELF Cutoffs Versus Experimentally Derived VCTE and FIB-4 Thresholds					
≥12.80	47.4% (24.4%, 71.1%)	≥75.0	33.3% (11.8%, 61.6%)	≥7.58	30.8% (14.3%, 51.8%)
≥12.00	21.3% (13.5%, 30.9%)	≥47.9	21.3% (12.7%, 32.3%)	≥5.47	21.3% (13.8%, 30.9%)
≥11.30	12.9% (9.1%, 17.4%)	≥29.5	12.9% (9.1%, 17.6%)	≥3.58	12.9% (9.3%, 17.2%)
≥10.50	7.2% (5.3%, 9.5%)	≥18.5	7.2% (5.3%, 9.5%)	≥2.24	7.2% (5.4%, 9.3%)
≥9.80	4.9% (3.7%, 6.4%)	≥12.1	4.9% (3.7%, 6.3%)	≥1.41	4.9% (3.7%, 6.3%)
≥9.00	4.1% (3.1%, 5.3%)	≥8.9	4.1% (3.1%, 5.3%)	≥0.96	4.1% (3.1%, 5.2%)
<9.00	0.0% (0.0%, 2.9%)	<8.9	1.2% (0.1%, 4.2%)	<0.96	0.0% (0.0%, 3.3%)
[a] Cutoff derived by interpolation from logistic regression.					

Gee, Sanyal et al- MS under review, presented at AASLD 202

Change in ELF score is associated with differential LRE risk

SIEMENS Healthineers

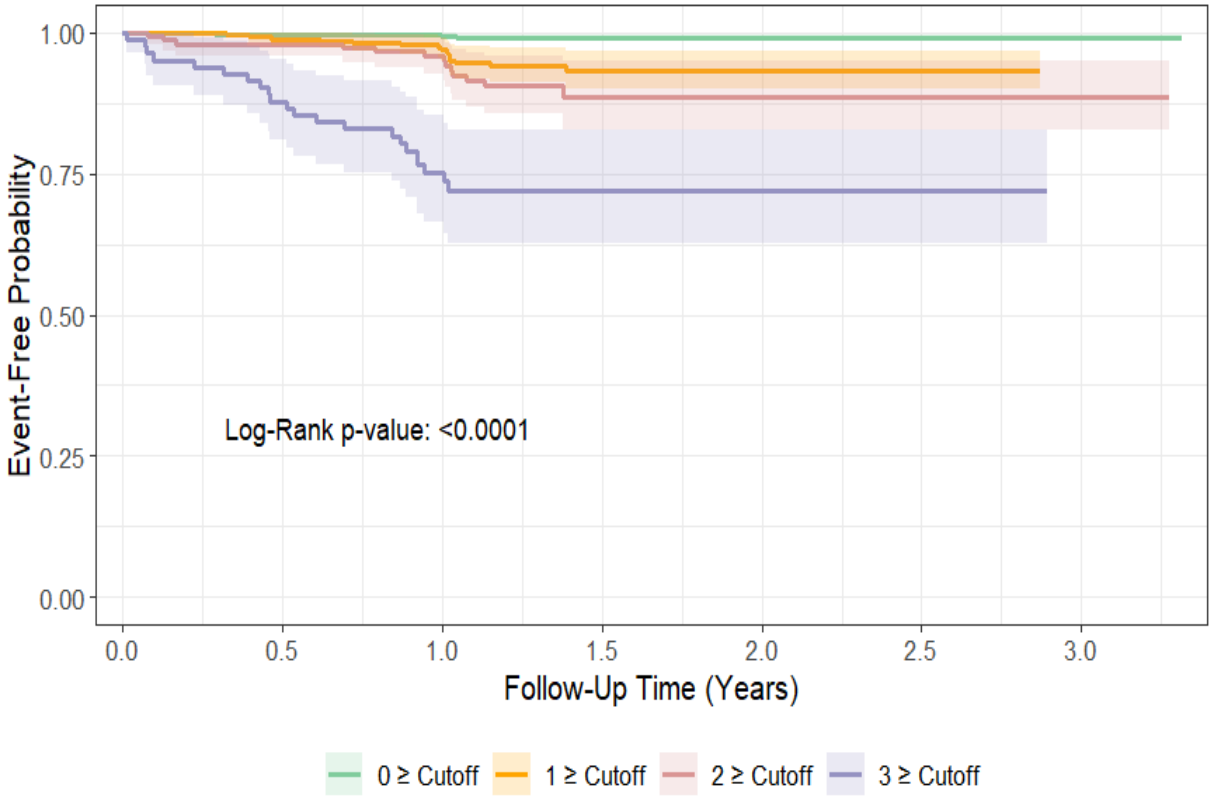
2071 Patients with F3-F4 MASH with ELF measurement at baseline and 1 year

ELF Score (Baseline)	ELF Change at 1 Year	N	Patients with LRE	LRE Risk (95% CI)	Interval Likelihood Ratio (95% CI)	LREs per 1000 Patient Years (95% CI)
≥11.30	Δ ≥0.5	81	13	16.0% (8.8%, 25.9%)	4.31 (2.48, 7.49)	124.0 (57.2, 200.2)
	Δ >−0.5 to <0.5	201	26	12.9% (8.6%, 18.4%)	3.35 (2.35, 4.76)	98.8 (64.6, 136.7)
	Δ ≤−0.5	108	7	6.5% (2.6%, 12.9%)	1.56 (0.75, 3.26)	40.4 (11.5, 75.0)
≥9.80 to <11.30	Δ ≥0.5	284	18	6.3% (3.8%, 9.8%)	1.52 (0.99, 2.34)	42.3 (23.5, 63.4)
	Δ >−0.5 to <0.5	619	14	2.3% (1.2%, 3.8%)	0.52 (0.32, 0.85)	15.3 (7.7, 24.1)
	Δ ≤−0.5	169	2	1.2% (0.1%, 4.2%)	0.27 (0.07, 1.07)	7.5 (0.0, 18.7)
<9.80	Δ ≥0.5	225	6	2.7% (1.0%, 5.7%)	0.62 (0.28, 1.35)	17.4 (5.8, 32.0)
	Δ >−0.5 to <0.5	317	1	0.3% (0.0%, 1.7%)	0.07 (0.01, 0.50)	1.9 (0.0, 5.7)
	Δ ≤−0.5	67	1	1.5% (0.0%, 8.0%)	0.34 (0.05, 2.43)	8.8 (0.0, 26.3)

Patients can be stratified by ELF change at 1 year: declined (≤ -0.5), stable (> -0.5 to < 0.5), increased (≥ 0.5)

Association of LRE risk with ELF change at 1 year is dependent on baseline ELF score

NIT concordance increases prognostic performance



Group	At Risk (Events)						
0	988 (0)	964 (2)	858 (5)	296 (7)	48 (7)	24 (7)	6 (7)
1	291 (0)	282 (3)	213 (8)	77 (15)	7 (15)	3 (15)	0 (15)
2	152 (0)	147 (3)	108 (6)	37 (13)	4 (13)	1 (13)	1 (13)
3	84 (0)	72 (10)	48 (20)	11 (22)	3 (22)	3 (22)	0 (22)

Fig. 2. Kaplan-Meier curves for event-free probability of liver-related outcomes. Risk stratification based on the number of NITs at or above cutoffs of ELF ≥11.30, VCTE ≥30.0 kPa and FIB-4 ≥3.48. Group 0 (Green): All 3 NITs below cutoffs; Group 1 (Orange): 1 NIT at or above cutoff; Group 2 (Red): 2 NITs at or above cutoffs; Group 3 (Purple): All 3 NITs at or above cutoffs

NIT criteria for enrollment:

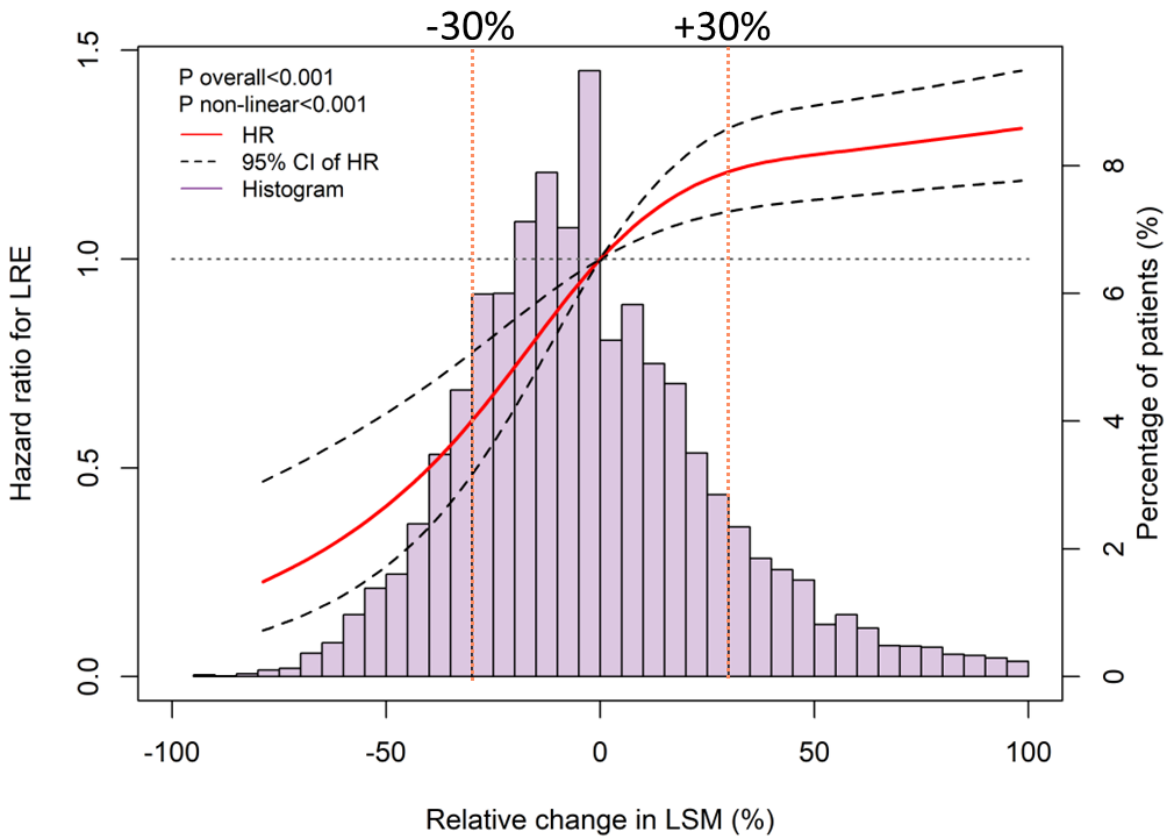
- Evidence of MASLD- MRI-PDFF > 5% or CAP > 280 db/m
- 3 NIT concordant for being above cutoff
  - FIB-4: 1- 2.6
  - LSM by VCTE: 9-20 Kpa
  - ELF: 9- 10.5 Kpa

Gee, Sanyal et al- MS under review, presented at AASLD 2024

In absence of treatment response vs outcomes data, how to leverage NIT prognostic and disease monitoring data to generate a RLSE?

Measuring change- VCTE LSM

Prospective study 16,603 patients, median follow-up 51.7 months
316 (1.9%) liver-related mortality/LRE outcomes.



First VCTE	% change	% of patients	LRE per 1,000 py
Low risk (LSM <10 kPa) N=9,009	>30% ↓	10%	1.4
	Stable	75%	0.7
	>30% ↑	15%	2.4
Intermediate risk (LSM 10-15 kPa) N=1,120	>30% ↓	37%	0.7
	Stable	52%	7.4
	>30% ↑	11%	17.2
High risk (LSM ≥15 kPa) N=791	>30% ↓	45%	16.1
	Stable	45%	33.3
	>30% ↑	10%	55.9

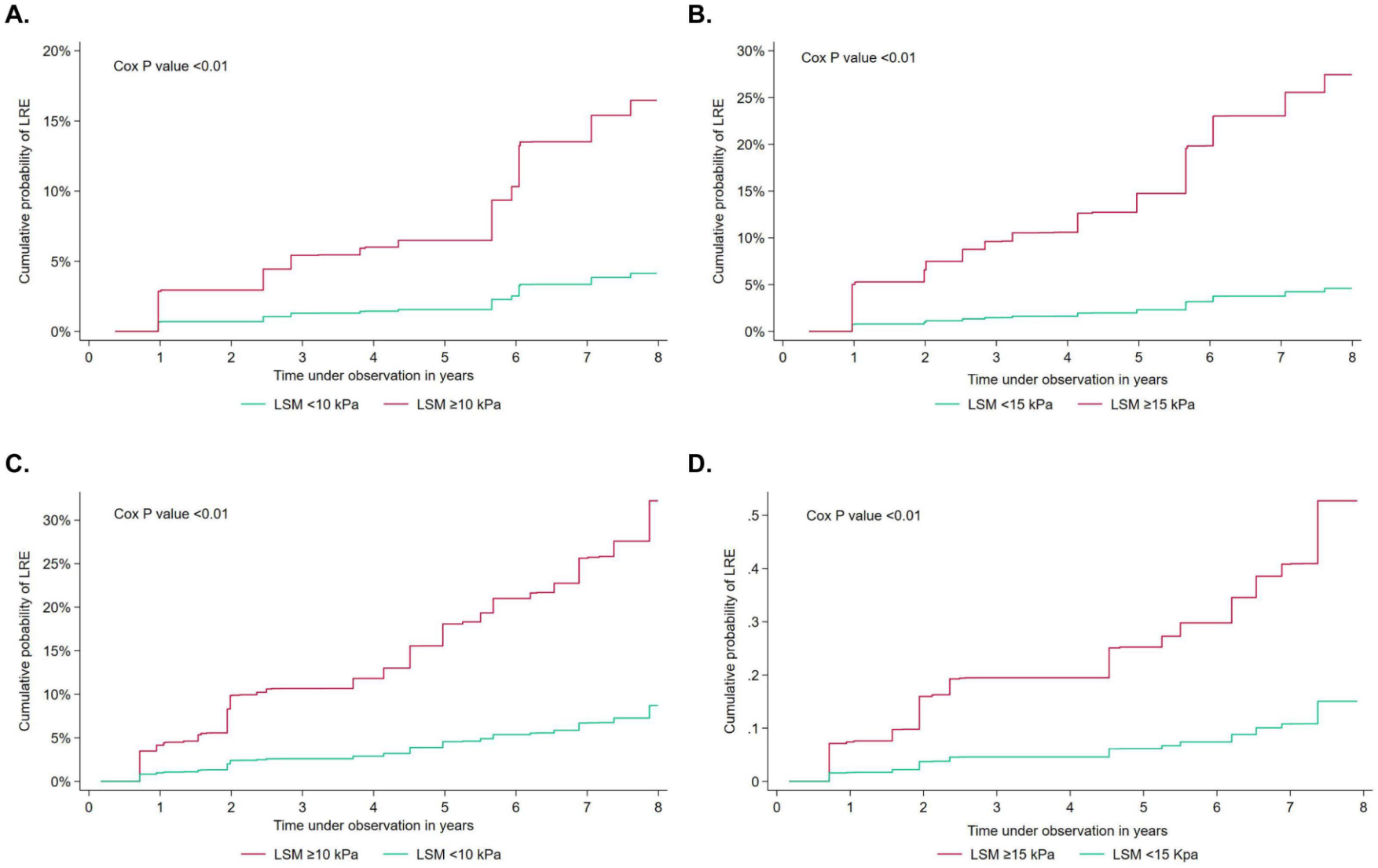
Lin et al. JAMA 2024;331:1287-1297

What's the problem then?

- Apply regulatory filters:
  - Target population- 10-20 Kpa
  - Follow up exam within time frame of typical phase 3 program (12-18 months)
  - The count outcomes

From starting point of 16K study participants and 316 outcomes to a very small number of outcomes

LSM change vs probability of outcome



Gawrieh et al, J Hepatol 2024

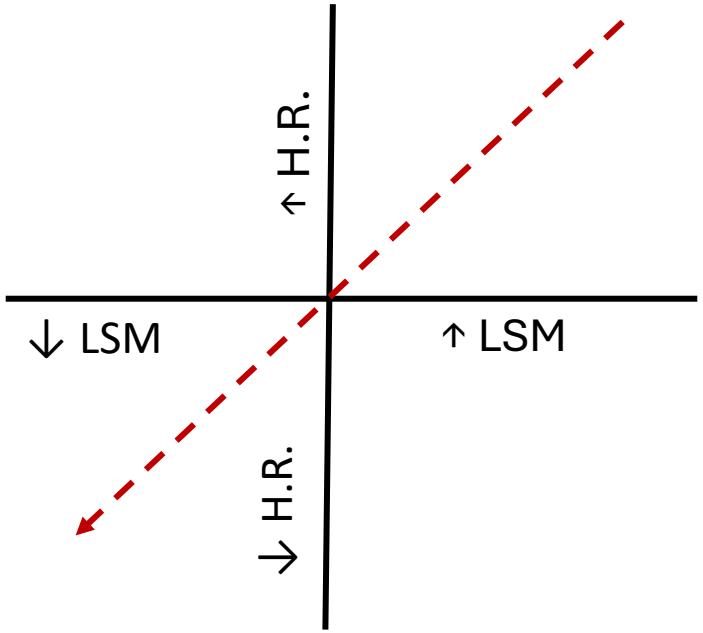
Limitations of sample size and outcome rates make it difficult to plot the clinical relevance of changes in LSM in the COU of a RLSE

	Univariate analysis		Multivariate analysis	
	HR (95% CI)	P value	HR (95% CI)	P value
	Baseline LSM > 10 Kpa			
↑ LSM 30%	1.73 (1-3.1)	0.05	1.75 (1-3.2)	0.05
↑ LSM 50%	1.97 (1.1-3.6)	0.01	1.83 (1-3.5)	0.05
	Baseline LSM > 20 Kpa			
↑ LSM 30%	1.4 (0.68-2.96)	n.s.	1.3 (0.6-2.76)	n.s.
↑ LSM 50%	1.9 (0.8-4.1)	n.s.	1.6 (0.7-3.4)	n.s.

Gawrieh et al, J Hepatol 2024

A road map to navigate regulatory landmines

Pooled large observational cohort
LSM 10-20 Kpa at baseline
f/U exam at 12-18 M
Regulatory outcomes included only

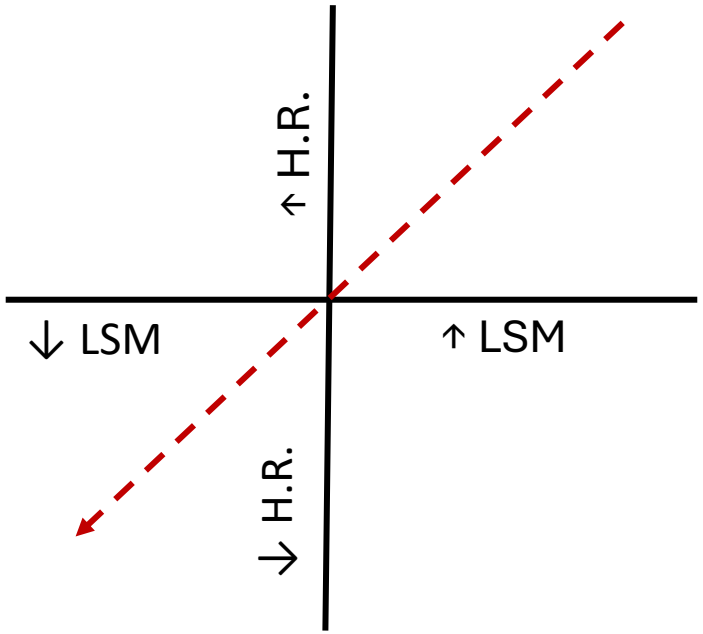


Cox model- time dependent

- Set LSM (or NIT) change above technical performance limit
- Establish change in H.R. is clinically meaningful and acceptable to both regulatory agencies and 3rd party payers
- Compute % change in NIT (LSM) that meets both criteria

A road map to navigate regulatory landmines

Assume a linear relationship

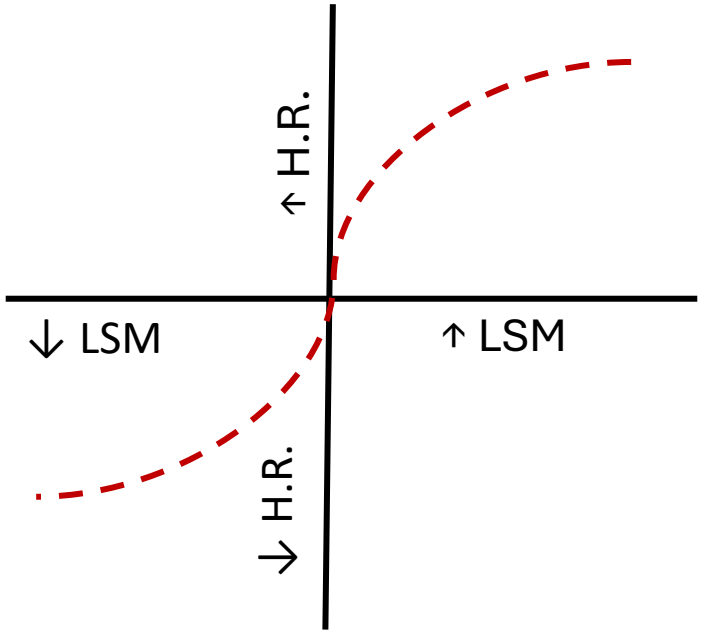


Assume change in LSM has same impact on HR in placebo and Rx arm

- If imputed HR change based on difference in % change in NIT (LSM) across arms is same across the dynamic range of measures
- Then it is valid to measure

A road map to navigate regulatory landmines

Assume a non-linear relationship



Assume change in LSM has same impact on HR in placebo and Rx arm

- If imputed HR change based on difference in % change in NIT (LSM) across arms is NOT the same across the dynamic range of measures
- May look at proportions in placebo vs active arms meeting a threshold considered clinically meaningful

Progress towards a RLSE- a report card

FDA accepts NIT based inclusion- thresholds at discretion of sponsor

Question	Key consideration
Which patients	How to define population at enough risk to warrant Rx and to permit demonstration of change in NIT that will translate in to clinical benefit?
NIT properties	1. Robustness of measure (analytical characteristics) √ 2. Prognostic performance √ 3. Sensitivity to change and time frame over which it changes- more data needed 4. How best to measure change- single vs multiple NITs, endpoint construction 5. Linking change to outcome risk- more work needed
Other characteristics	Can it be deployed at scale? Is it translatable? Cost?
Single vs multiple NITs	Need to be certain enough to claim “reasonably likely”

Type of evidence burden is same for all NITs to be considered a surrogate outcome and RLSE

Thank you for your attention

When in doubt keep climbing