



Published in final edited form as:

*J Hepatol.* 2023 September ; 79(3): 592–604. doi:10.1016/j.jhep.2023.04.025.

## An individual patient data meta-analysis to determine cut-offs for and confounders of NAFLD-fibrosis staging with magnetic resonance elastography

Jia-xu Liang<sup>1,2,3</sup>, Javier Ampuero<sup>1,2</sup>, Hao Niu<sup>4,5</sup>, Kento Imajo<sup>6</sup>, Mazen Nouredin<sup>7</sup>, Jaideep Behari<sup>8</sup>, Dae Ho Lee<sup>9</sup>, Richard L. Ehman<sup>10</sup>, Fredrik Rorsman<sup>11</sup>, Johan Vessby<sup>11</sup>, Juan R. Lacalle<sup>12</sup>, Ferenc E. Mózes<sup>13</sup>, Michael Pavlides<sup>13,14</sup>, Quentin M. Anstee<sup>15,16</sup>, Stephen A. Harrison<sup>13</sup>, Javier Castell<sup>17</sup>, Rohit Loomba<sup>18,19</sup>, Manuel Romero-Gómez<sup>1,2,\*</sup> LITMUS Consortium Investigators

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

\*Corresponding author. Address: Digestive Disease Department, Virgen del Rocío University Hospital, Avenida Manuel Siurot s/n, 41013, Sevilla, Spain. Tel.: +34 954 786 587. mromerogomez@us.es (M. Romero-Gómez). <https://doi.org/10.1016/j.jhep.2023.04.025>.

Authors' contributions

Guarantor of the article: MR-G. Study design, statistical analyses, and interpretation: MR-G, JxL, JA, HN. Drafting the manuscript: JxL, JA, MR-G. Data acquisition and critical review of the manuscript: all authors. Approved the final version of the article, including the authorship list: all authors.

Conflicts of interest

The Mayo Clinic and RLE have intellectual property rights and a financial interest in MRE technology. RL serves as a consultant to Aardvark Therapeutics, Altimmune, Anylam/Regeneron, Amgen, Arrowhead, AstraZeneca, Bristol Myer Squibb, CohBar, Eli Lilly, Galmed, Gilead, Glympse Bio, HighTide, Inipharma, Intercept, Inventiva, Ionis, Janssen, Madrigal, Metacrine, NGM Bio, Novartis, Novo Nordisk, Merck, Pfizer, Sagimet, Theratechnologies, 89bio, Terns, and Viking. In addition, his institutions have received research grants from Arrowhead, AstraZeneca, Boehringer Ingelheim, Bristol Myers Squibb, Eli Lilly, Galectin, Galmed, Gilead, Intercept, Hanmi, Intercept, Inventiva, Ionis, Janssen, Madrigal, Merck, NGM Bio, Novo Nordisk, Merck, Pfizer, Sonic Incytes, and Terns. He is also a co-founder of LipoNexus Inc. QMA is a coordinator of the EU IMI-2 LITMUS consortium, which is funded by the EU Horizon 2020 programme and EFPIA. His institutions have received research grants from Abbvie, AstraZeneca, Boehringer Ingelheim, Glympse Bio, Intercept, Novartis, Pfizer, Allergan/Tobira, GlaxoSmithKline, Glympse Bio, Intercept, and Novartis. He also serves as a consultant on behalf of Newcastle University for Alimentiv, Akero, AstraZeneca, Axcella, 89bio, Boehringer Ingelheim, Bristol Myers Squibb, Galmed, Genfit, Genentech, Gilead, GlaxoSmithKline, Hanmi, HistoIndex, Intercept, Inventiva, Ionis, IQVIA, Janssen, Madrigal, Medpace, Merck, NGM Bio, Novartis, Novo Nordisk, PathAI, Pfizer, Poxel, Resolution Therapeutics, Roche, Ridgeline Therapeutics, RTI, Shionogi, and Terns. In addition, he is a speaker for Fishawack, Integritas Communications, Kenes, Novo Nordisk, Madrigal, Medscape, and Springer Healthcare, and receives royalties from Elsevier Ltd. MP is a shareholder in Perspectum Ltd. MN has been on the advisory board and served as a consultant for 89bio, Altimmune, Gilead, CohBar, Cytodyn, ChronWell, Intercept, Pfizer, Novo Nordisk, Blade, EchoSens, Fractyl, Madrgial, NorthSea, Prespectum, Terns, Siemens, and Roche Diagnostics. Moreover, he has received research support from Allergan, BMS, Gilead, Galmed, Galectin, Genfit, Conatus, Enanta, Madrigal, Novartis, Pfizer, Shire, Viking, and Zydus. He is also a shareholder or has stocks in Anaetos, Ciema, Rivus Pharma, and Viking. SAH serves as a scientific advisor or consultant for Akero, Alentis, Altimmune, Arrowhead, Axcella, BMS, ChronWell, Echosens, Galectin, Gilead, Hepion, Hepagene, HistoIndex, Intercept, Madrigal, Medpace, Metacrine, NGM Bio, NorthSea, Novartis, Novo Nordisk, PathAI, Poxel, Sagimet, Terns, and Viking. He has stock options in Akero, Cirus, Galectin, Genfit, Hepion, HistoIndex, PathAI, Metacrine, NGM Bio, and NorthSea. In addition, he has received grants and/or research support from Akero, Axcella, BMS, Cirus, CiVi Biopharma, Conatus, Corcept, Cymabay, Enyo, Galectin, Genentech, Genfit, Gilead, Hepion, HighTide, Intercept, Madrigal, Metacrine, NGM Bio, Novartis, Novo Nordisk, NorthSea, Pfizer, Sagimet, Viking, and 89bio. FR is a member of the advisory board for Norgine and Intercept; has received a speaker fee from Norgine and Gore; and has received research support from Norgine, Antaros Medical, and Boehringer Ingelheim. JV has received research support from Antaros Medical and Boehringer Ingelheim. JB has received research grant funding from Gilead, Pfizer, and ENDRA Life Sciences. In addition, he has signed institutional research contracts with Intercept, Pfizer, Galectin, Exact Sciences, Inventiva, Enanta, Shire, Gilead, Allergan, Celgene, Galmed, and Genentech. MR-G is a scientific advisor, consultant, and/or speaker for Abbvie, Alpha Sigma, Allergan, AstraZeneca, Axcella, BMS, Boehringer Ingelheim, Gilead, Intercept, Inventia, Kaleido, MSD, Novo Nordisk, Pfizer, Prosciento, Rubió, Siemens, Shionogi, Sobi, and Zydus. He has received research grants from Gilead, Intercept, and Siemens, and is a co-inventor of Hepamet Fibrosis Score, DeMILI, and DeMILI 3.0. The other authors declare no conflict of interest that pertain to this work. Please refer to the accompanying ICMJE disclosure forms for further details.

Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jhep.2023.04.025>.

- <sup>1</sup>Digestive Diseases Unit and CIBERehd, Virgen del Rocío University Hospital, Seville, Spain
- <sup>2</sup>Institute of Biomedicine of Seville (HUVR/CSIC/US), University of Seville, Seville, Spain
- <sup>3</sup>Department of Diagnostic Radiology, The Fifth Clinical Medical College of Henan University of Chinese Medicine (Zhengzhou People's Hospital), Zhengzhou, China
- <sup>4</sup>Digestive System and Clinical Pharmacology Unit, Virgen de la Victoria University Hospital, Biomedical Research Institute of Malaga and Nanomedicine Platform-IBIMA (Plataforma BIONAND), University of Malaga, Málaga, Spain
- <sup>5</sup>Biomedical Research Network Center for Hepatic and Digestive Diseases (CIBERehd), Carlos III Health Institute, Madrid, Spain
- <sup>6</sup>Department of Gastroenterology, Yokohama City University Graduate School of Medicine; Yokohama, Japan
- <sup>7</sup>Fatty Liver Program, Division of Digestive and Liver Diseases, Comprehensive Transplant Program, Cedars-Sinai Medical Center, Los Angeles, CA, USA
- <sup>8</sup>Department of Medicine, Division of Gastroenterology, Hepatology and Nutrition, Center for Liver Diseases, University of Pittsburgh Medical Center, Pittsburgh, PA, USA
- <sup>9</sup>Department of Internal Medicine, Gachon University College of Medicine (Gachon University Gil Medical Center), Incheon, South Korea
- <sup>10</sup>Department of Diagnostic Radiology, Mayo Clinic College of Medicine, Rochester, MN, USA
- <sup>11</sup>Department of Medical Sciences, Section of Gastroenterology and Hepatology, Uppsala University, Uppsala, Sweden
- <sup>12</sup>Biostatistics Unit, Department of Preventive Medicine and Public Health, University of Seville, Seville, Spain
- <sup>13</sup>Oxford Centre for Clinical Magnetic Resonance Research, Division of Cardiovascular Medicine, Radcliffe Department of Medicine, University of Oxford, Oxford, UK
- <sup>14</sup>Translational Gastroenterology Unit, University of Oxford, Oxford, UK
- <sup>15</sup>Translational and Clinical Research Institute; Faculty of Medical Sciences, Newcastle University, Newcastle upon Tyne, UK
- <sup>16</sup>Newcastle NIHR Biomedical Research Centre, Newcastle upon Tyne Hospitals, NHS Trust, Newcastle upon Tyne, UK
- <sup>17</sup>Department of Radiology, Virgen del Rocío University Hospital, Seville, Spain
- <sup>18</sup>Division of Epidemiology, Department of Family Medicine and Public Health, University of California San Diego, La Jolla, CA, USA
- <sup>19</sup>NAFLD Research Center, Division of Gastroenterology and Hepatology, Department of Medicine, University of California San Diego, La Jolla, CA, USA

## Abstract

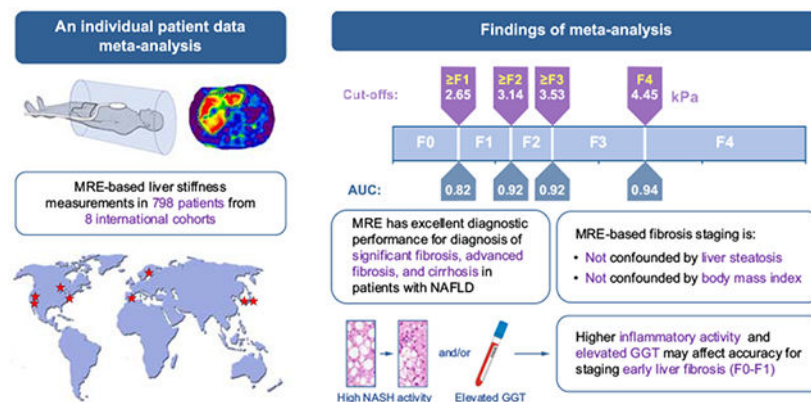
**Background & Aims:** We conducted an individual patient data meta-analysis to establish stiffness cut-off values for magnetic resonance elastography (MRE) in staging liver fibrosis and to assess potential confounding factors.

**Methods:** A systematic review of the literature identified studies reporting MRE data in patients with NAFLD. Data were obtained from the corresponding authors. The pooled diagnostic cut-off value for the various fibrosis stages was determined in a two-stage meta-analysis. Multilevel modelling methods were used to analyse potential confounding factors influencing the diagnostic accuracy of MRE in staging liver fibrosis.

**Results:** Eight independent cohorts comprising 798 patients were included in the meta-analysis. The area under the receiver operating characteristic curve (AUROC) for MRE in detecting significant fibrosis was 0.92 (sensitivity, 79%; specificity, 89%). For advanced fibrosis, the AUROC was 0.92 (sensitivity, 87%; specificity, 88%). For cirrhosis, the AUROC was 0.94 (sensitivity, 88%, specificity, 89%). Cut-offs were defined to explore concordance between MRE and histopathology: F2, 3.14 kPa (pretest probability, 39.4%); F3, 3.53 kPa (pretest probability, 24.1%); and F4, 4.45 kPa (pretest probability, 8.7%). In generalized linear mixed model analysis, histological steatohepatitis with higher inflammatory activity (odds ratio 2.448, 95% CI 1.180–5.079,  $p < 0.05$ ) and high gamma-glutamyl transferase (GGT) concentration ( $>120\text{U/L}$ ) (odds ratio 3.388, 95% CI 1.577–7.278,  $p < 0.01$ ) were significantly associated with elevated liver stiffness, and thus affecting accuracy in staging early fibrosis (F0–F1). Steatosis, as measured by magnetic resonance imaging proton density fat fraction, and body mass index (BMI) were not confounders.

**Conclusions:** MRE has excellent diagnostic performance for significant, advanced fibrosis and cirrhosis in patients with NAFLD. Elevated inflammatory activity and GGT level may lead to overestimation of early liver fibrosis, but anthropometric measures such as BMI or the degree of steatosis do not.

## Graphical abstract



## Keywords

Elastography; Magnetic Resonance; Fibrosis; NAFLD; NASH; GGT

## Introduction

NAFLD is the leading cause of chronic liver disease worldwide, affecting approximately 25% of the global population.<sup>1</sup> It is a spectrum comprising isolated steatosis, steatohepatitis, advanced fibrosis, and cirrhosis.<sup>2</sup> Liver fibrosis predicts the occurrence of liver-related events and overall mortality in patients with NAFLD.<sup>3</sup> Therefore, the accurate detection and staging of liver fibrosis is a key issue for the management of patients with NAFLD.<sup>4</sup>

Liver biopsy is the reference standard for assessing the liver fibrosis stage in patients with NAFLD. However, it cannot be widely used in clinical practice owing to its limitations and risks, such as sampling errors, poor tolerance, its high cost, and the risk of serious adverse events.<sup>5,6</sup> Therefore, several biochemical<sup>7,8</sup> and imaging biomarkers have emerged to help in the diagnose of liver fibrosis, with imaging mechanical biomarkers such as transient elastography (TE), shear-wave elastography (SWE), and magnetic resonance elastography (MRE) playing a major role among them. MRE evaluates a larger portion of the liver, offering the option of choosing the region of interest, which can be performed in conjunction with conventional magnetic resonance imaging.<sup>9</sup> Several studies reported that MRE has a higher diagnostic performance than TE and SWE in assessing liver fibrosis in patients with NAFLD,<sup>10-13</sup> including morbidly obese patients and those with ascites, additionally decreasing the failure rate. However, there are cases in which a discrepancy between fibrosis based on histology and liver stiffness measurement (LSM) by MRE occurs. In addition, there is not a generally accepted cut-off value for diagnosing the different stages of fibrosis because most of the published studies have been based on small and heterogenous cohorts. Although two individual patient data meta-analyses (IPD-MAs)<sup>14,15</sup> have been previously published and presented their respective cut-off values for staging liver fibrosis, they have not been directly compared and validated in a large sample cohort. The aim of this study was therefore to conduct an IPD-MA including data from studies that evaluated LSM by MRE with histology as the reference standard, to establish diagnostic cut-offs and explore the potential confounding factors influencing the diagnostic accuracy of MRE for staging liver fibrosis in patients with NAFLD.

## Materials and methods

Our study was an IPD-MA based on a multicentre collaboration from previously published studies. It was performed in accordance with the PRISMA-IPD Statement checklist<sup>16</sup> and PRISMA-DTA Statement checklist,<sup>17</sup> and it was registered in PROSPERO (CRD42021264458).

### Search strategy and selection criteria

Two investigators (JxL and JA) independently identified suitable studies, published up to 31 December 2021, by searching in the MEDLINE (using PubMed as the search engine), Embase, and Cochrane databases; discrepancies were solved by a third author (MR-G). A combination of keywords and medical subject heading terms were used, including ('MR' OR 'Magnetic resonance') AND ('Elastography' OR 'Elasticity' OR 'MRE') AND ('NAFLD' OR 'Non-alcoholic fatty liver disease' OR 'MAFLD' OR 'Metabolic associated fatty liver disease'). A manual search was conducted by using references listed in the

original articles and review articles retrieved. The search was restricted to English-language literature. The inclusion criteria were as follows: (1) patients older than 18 years; (2) no presence of other comorbid liver diseases; (3) assessment of liver fibrosis by using MRE; (4) histopathology as the reference standard; and (5) access to individual patient data. The exclusion criteria were as follows: (1) duplicate reports; (2) studies conducted on animals; (3) systematic reviews or meta-analyses; and (4) insufficient data to perform the IPD-MA despite having contacted the study investigators.

Once relevant studies were identified, the corresponding author of the eligible study was contacted and provided with the background and details of the individual patient data (IPD) pooled analysis, as well as a data collection file for the input of individual patient results for the project. Then, the completed IPD collection files were sent back to the data management team. In case of non-response, the study was excluded after two contact attempts. Data harmonisation was performed by at least two independent readers (JxL, JA, JRL, and MR-G) at the site of the central data management team. Discrepancies, missing data, obvious errors, and inconsistencies between variables or outlying values were queried and rectified as necessary with input from the original authors. Finally, the following patient records were excluded: (1) no MRE measurements reported or MRE technical failure; (2) lack of essential clinical information; and (3) lack of relevant histopathological data.

### Data abstraction

Reviewers extracted data and assessed the quality of each study. Differences in opinion between reviewers were resolved by consensus with a third reviewer. The following variables were recorded.

1. Study characteristics: study location, design, publication year, time period of study, interval between MRE and biopsy, and number of patients with NAFLD.
2. Demographic data: age, sex, BMI, weight, type 2 diabetes mellitus (T2DM), and arterial hypertension.
3. Laboratory test: aspartate aminotransferase (U/L), alanine aminotransferase (U/L), gamma-glutamyl transferase (GGT) (U/L), albumin (g/dl), fasting blood glucose (mg/dl), triglyceride (mg/dl), total cholesterol (mg/dl), platelet count ( $10^9/L$ ), HDL cholesterol (mg/dl), and LDL cholesterol (mg/dl).
4. Imaging data: LSM by MRE. All MREs were performed with 3-T/1.5-T field strength scanners using spin-echo echo-planar imaging (SE-EPI)/gradient-echo sequence (GRE) sequences, and shear waves at 60 Hz. To quantify the liver fat content, a multi-echo gradient-echo sequence was used to obtain magnetic resonance imaging (MRI) proton density fat fraction from a single breath-hold acquisition; proton density fat fraction was defined to be the proportion of magnetic resonance (MR)-visible fat protons to the sum of MR-visible fat and water protons.  $R2^*$  maps (relaxation rate =  $1/T2^*$ ) data were also collected if performed.
5. Histopathological data: Histopathological data were extracted according to the Brunt scoring system ( $n = 58$ ),<sup>18</sup> the NASH–Clinical Research Network (CRN)

( $n = 617$ ),<sup>19,20</sup> and the SAF (steatosis–activity–fibrosis) score ( $n = 123$ ).<sup>21</sup> Steatosis was defined according to the number of affected hepatocytes: S0 (<5% depending on the slide), S1 (5–33%), S2 (34–66%), and S3 (>66%). In the SAF histological scoring system,<sup>21</sup> each of the three features (steatosis, hepatocyte ballooning, and lobular inflammation) classified as at least grade 1 was defined as NASH, and activity (lobular inflammation 0–2 + hepatocyte ballooning 0–2) 3 was classified as severe activity. In the NASH-CRN scoring system, the NAFLD activity score (NAS) ranged from 0 to 8 and consisted of steatosis (0–3), lobular inflammation (0–3), and hepatocyte ballooning (0–2). Therefore, NAS 4, with a subscore of 1 or higher for each subcomponent, was defined as NASH in this scoring system; lobular inflammation (0–3) + hepatocyte ballooning (0–2) 3 was also classified as severe activity to avoid the effect of the degree of steatosis. Moreover, on the basis of differentiating between NASH and no-NASH, all patients with NASH were classified as either NASH mild–moderate activity (NASH-MMA) or NASH severe activity (NASH-SA). These classification processes are shown in Fig. S1. Liver fibrosis was scored stage 0 (F0; none), stage 1 (F1; 1a or 1b perisinusoidal zone 3 or 1c portal fibrosis), stage 2 (F2; perisinusoidal and periportal fibrosis without bridging), stage 3 (F3; bridging fibrosis), and stage 4 (F4; cirrhosis).

### Quality assessment

The risk of bias and applicability concerns of the included cohorts were assessed by two independent readers using the QUADAS-2 tool.<sup>22</sup> The quality assessment consisted of the evaluation of four components: patient selection, index tests, reference standard, and flow and timing (Fig. S2).

### Statistical analysis

All statistical analyses were performed using MedCalc (version 19.7, MedCalc Software Inc, Ostend, Belgium) and Stata 16 (StataCorp LP, College Station, TX, USA). All statistical tests were two-sided with a significance level of 5%. For the overall population, continuous variables were reported as means with SDs, or medians with IQRs according to the data distribution. Categorical variables were displayed as percentages.

Owing to the unbalanced distribution of liver fibrosis, we performed a two-stage method in the pooled cut-off meta-analysis. First, we replicated 20,000 stratified bootstrap samples to estimate the optimal cut-off (based on the Youden index)<sup>23</sup> and 95% CI of each fibrosis stage in an independent cohort. Second, we used these study-level cut-off values to estimate pooled cut-offs in a fixed- or random-effects model meta-analysis. Random-effects estimation was carried out if heterogeneity was higher than 50%; fixed-effects analysis was used whenever it was lower, and the pooled stiffness cut-off values were then displayed graphically as forest plots. Afterwards, the validation group generated by 3:1 random sampling was used to compare our pooled cut-off values with those of the previous IPD-MAs.<sup>14,15</sup> True-positive, false-positive, true-negative, and false-negative rates were calculated by sensitivity and specificity (Youden index) for each included cohort. Summary sensitivity, specificity, positive likelihood ratio, negative likelihood ratio,

diagnostic odds ratio, and area under the receiver operating characteristic curve (AUROC) with corresponding 95% CIs were calculated using the bivariate random-effects model to examine the diagnostic accuracy of MRE.<sup>24</sup> Heterogeneity was assessed using the Higgins inconsistency index ( $I^2$ ) test, with values greater than 50% taken as an indicator of substantial heterogeneity.<sup>25</sup>

Several preplanned subgroups and stratified analyses were performed based on study population (Asian or non-Asian), sex (male or female), T2DM (presence or absence), arterial hypertension (presence or absence), presence or absence of obesity (BMI  $\geq 30$  kg/m<sup>2</sup> or BMI  $<30$  kg/m<sup>2</sup>), presence or absence of NASH, liver steatosis degree (steatosis 0–1 or 2–3), sequence (GRE or SE-EPI), field strength (1.5 T or 3.0 T), and manufacturer (Siemens and Philips or GE) of MRI. Subgroup AUCs were compared using the method by DeLong *et al.*<sup>26</sup> Furthermore, a sensitivity analysis was also performed to analyse the effect of the interval between MRE and liver biopsy, and R2\* on the diagnostic performance of MRE. Univariate meta-regression analysis was also used to explore possible sources of heterogeneity; the covariates of this study level included the following: (1) sequence (GRE or SE-EPI), (2) field strength (1.5 T or 3.0 T), (3) manufacturer (Siemens and Philips or GE) of MRI, (4) study design (prospective or retrospective), and (5) population (Asian or non-Asian). We assessed the potential publication bias using Deeks' funnel plot asymmetry test, in which  $p < 0.10$  indicated statistical significance.

To assess factors affecting the discordance between MRE-based staging and histopathology, multilevel models were fitted, using a one-stage approach.<sup>27</sup> Linear mixed-effects model (LMM) and generalised linear mixed model (GLMM), as extensions of traditional linear and logistic regression, were often used when heterogeneity was present and IPD were available. Some continuous missing data were replaced by the mean or median value according to the data distribution. A previously published study,<sup>28</sup> which lacked too much data, was not included in this part analysis. LMM was used to explore the effect of covariates as fixed effects on MRE, with different research centres as random effects. There are two categories of significant discordance: overestimation and underestimation. Overestimation was defined as the stage of MRE determination according to the previous pooled cut-off values at least two stages above histopathology or histopathological findings of F3 with LSM by MRE greater than 7.5 kPa. Underestimation was defined as the stage of MRE determination according to the previous pooled cut-off at least two stages lower than histopathology. For these binary outcomes (concordance/overestimation and concordance/underestimation), a GLMM was chosen to explore the covariates associated with discrepancies between LSM by MRE and histological assessment of fibrosis grading. Covariates with  $p < 0.10$  in the univariate analysis were entered into the multivariate analysis. Finally, the relationship between the selected covariates in GLMM and LSM by MRE was assessed using two-way ANOVA.

In additional analyses, 'rule-in' (at a specificity of 90%) and 'rule-out' (at a sensitivity of 90%) cut-off values for LSM by MRE were identified with a traditional receiver operating characteristic analysis method. Based on Bayes' theorem,<sup>29</sup> we further investigated the impact of different prevalence settings on the predictive value under the three cut-off values (pooled cut-off, 'rule-in' cut-off, and 'rule-out' cut-off).

## Results

From 252 unique studies identified using our search strategy, 16 diagnostic studies<sup>10-12,28,30-41</sup> from nine authors were eligible. All nine authors were contacted; however, no response was received from three authors.<sup>30-32</sup> The remaining six authors agreed to share their data. Moreover, two additional independent cohorts from Virgen del Rocío University Hospital, Seville (n = 61),<sup>42</sup> and the Department of Gastroenterology and Hepatology at Uppsala University, Sweden (n = 66),<sup>43</sup> were included after the manual search. Fig. 1 shows the detailed study identification and selection flow chart. We updated the additional data of each included cohort in January 2022 and received the data of 821 patients from eight independent cohorts. After screening, 23 patients were excluded, leaving a total of 798 patients in our final IPD-MA.

### Baseline study and patient characteristics

The total analysis set comprised 798 patients from eight international patient cohorts. Four cohorts were from the USA (n = 451), two cohorts were from Europe (n = 123), and the remaining two cohorts were from Asia (n = 224). Two cohorts were retrospective studies, and the remaining cohorts were prospectively collected; two cohorts used 1.5-T MRI scanners, and six cohorts used 3.0-T MRI scanners. A 2D SE-EPI sequence was performed in four cohorts, and the remaining with a GRE sequence; all the studies had shear waves generated at 60 Hz. Five cohorts used the NASH-CRN scoring system as a reference standard. Two cohorts were based on the SAF scoring system, and one cohort was based on the Brunt scoring system. Baseline study characteristics are given in Table 1.

The mean age of the pooled cohort was  $51.6 \pm 14.8$  years, and 44% were male with a mean BMI of  $31.9 \pm 6.0$  kg/m<sup>2</sup>. The median interval between the performance of MRE and liver biopsy was 41 days (IQR 15–87.25 days). The mean liver stiffness across the entire cohort was  $3.26 \pm 1.51$  kPa, ranging from 1.2 to 10.9 kPa. The distribution of fibrosis in the pooled cohort was as follows: F0, 203 (25.4%); F1, 281 (35.2%); F2, 122 (15.3%); F3, 123 (15.4%); and F4, 69 (8.7%). Fig. 2 shows the mean liver stiffness values corresponding to each histological liver fibrosis stage. The distribution of steatosis was as follows: grade 0, 5.5%; grade 1, 42.7%; grade 2, 37.5%; and grade 3, 14.3%. According to the SAF and NASH-CRN scoring systems, 425 patients were defined as NASH. The remaining important clinical information is available in Table 2.

### Quality of included studies

The methodological quality of the studies assessed using the QUADAS-2 tool is summarised in Fig. S2A and B. The retrospective design of two cohorts introduced unclear risks in the patient selection domain. All studies did not report predefined cut-off values and were judged as having a high risk of bias in the index test domain. The reference standard domain was judged to have an unclear risk of bias in all included studies owing to the inherent limitations of liver biopsy.



### Pooled cut-offs and diagnostic accuracy of MRE

The diagnostic performance of MRE for staging fibrosis and heterogeneity ( $I^2$ ) is shown in Table 3, and the main pooled results are depicted in Fig. 3A-D and Fig. S3A-D.

For mild fibrosis ( F1), analysis was done on seven cohorts; one cohort was not included as there were not enough patients to calculate reliable cut-off values. The optimal cut-off value determined in the meta-analysis was 2.65 kPa (95% CI 2.52–2.78 kPa), with an  $I^2$  statistic of 17.05.

For significant fibrosis ( F2), analysis of the cut-off value was done with all eight cohorts. The optimal cut-off value determined in the meta-analysis was 3.14 kPa (95% CI 3.01–3.27 kPa), with an  $I^2$  statistic of 44.88.

For advanced fibrosis ( F3), analysis of the cut-off value was done with all eight cohorts. The optimal cut-off value determined in the meta-analysis was 3.53 kPa (95% CI 3.40–3.66 kPa), with an  $I^2$  statistic of 41.15.

For cirrhosis (F4), analysis was done on six cohorts; two cohorts were not included owing to the low prevalence of cirrhosis to calculate reliable cut-off values. The optimal cut-off value determined in the random-effects meta-analysis was 4.45 kPa (95% CI 3.63–5.27 kPa), with an  $I^2$  statistic of 85.24.

### Subgroup and sensitivity analysis

In subgroup analysis (Table S1), the AUROC (0.86, 95% CI 0.82–0.90) in patients with type 2 diabetes was significantly higher than the AUROC (0.75, 95% CI 0.70–0.79,  $p < 0.01$ ) in patients without type 2 diabetes for staging mild fibrosis. The presence of NASH, as defined by histopathological scoring systems, showed significantly lower diagnostic accuracy than did the absence of NASH in staging significant fibrosis (no-NASH 0.93 [0.90–0.96] vs. NASH 0.87 [0.83–0.90],  $p < 0.01$ ) and advanced fibrosis (no-NASH 0.96 [0.93–0.98] vs. NASH 0.90 [0.87–0.93],  $p < 0.01$ ). The diagnostic accuracy of MRE was not affected by other factors. Patients with an interval between MRE and liver biopsy  $>6$  and  $>3$  months and high  $R2^*$  ( $R2^* > 80$  Hz) were removed from the sensitivity analysis, and the diagnostic performance of MRE in the remaining patients was not significantly different compared with the entire IPD study group.

### Meta-regression

Univariate meta-regressions showed that these study-level variables (population, study design, field strength, sequence, and manufacturer of MRI) were not associated with heterogeneity (see Fig. S4A-D).

### Comparison of three cut-off value systems in the overall cohort

A random sampling group was used as a ‘validation set’ to avoid overlapping of data as often as possible. The diagnostic performance of these three IPD-MAs (our current study, the study by Singh *et al.*,<sup>14</sup> and the study by Hsu *et al.*<sup>15</sup>) for staging liver fibrosis is shown

in Table S2. However, there was no significant difference in accuracy between these three cut-off value systems.

### Factors associated with liver stiffness measured by MRE

Table 4 shows the results of the LMM to determine the influence of covariates on MRE values. Age, GGT concentration, platelet count, fibrosis stage, and NASH-SA were found to have significant and relevant relationships with LSM by MRE.

### Factors affecting diagnostic accuracy of MRE

In GLMM, we found that NASH-SA (odds ratio [OR] 2.448, 95% CI 1.180–5.079,  $p < 0.05$ ) and high GGT level ( $>120$  U/L) (OR 3.388, 95% CI 1.577–7.278,  $p < 0.01$ ) were independent factors causing overestimation (Table 5). Conducting a multivariate analysis excluding histological variables (which are difficult to obtain in clinical practice), high GGT level (OR 3.700, 95% CI 1.749–7.825,  $p < 0.01$ ) was the only significant variable in the multivariate analysis to be implemented in clinical practice (Table 6). No significant confounding factors were found between concordance and underestimation patients.

### Relationship between steatohepatitis presence, GGT levels, and LSM by MRE in every stage of liver fibrosis

The relationship between LSM by MRE, GGT concentration ( $<60$ ,  $60$ – $120$ , and  $>120$  U/L), and NASH levels (no-NASH, NASH-MMA, and NASH-SA) at each fibrosis stage (F0–F4) was analysed using two-way ANOVA. The results showed that NASH-SA and higher GGT levels increased LSM by MRE, but only significantly in patients with mild fibrosis stage (F0 and F1), and did not interact with each other (Table S3). Notably, the LSM in the no-NASH group (5.14 kPa, 95% CI 4.70–5.57 kPa) was higher than that in the NASH-MMA group (4.58 kPa, 95% CI 3.82–5.33 kPa) in advanced fibrosis (F3–F4), although not statistically significant. The LSM ranges of no-NASH/NASH-MMA and NASH-SA seemed to show a good distinction in the F0–F1 phase, although there was still a small overlap between them (Fig. S5).

### Publication bias

According to Deeks' funnel plot asymmetry test, publication bias was not present ( $p > 0.10$ ) (Fig. S6A–D).

### Additional analysis

The overall diagnostic accuracy of 'rule-out' and 'rule-in' cut-offs is presented in Table S4. To investigate the impact of fibrosis prevalence on the predictive value of LSM by MRE, we set out to calculate positive and negative predictive values by using a range of different prevalences. The prevalence figures were used to represent values from our cohort ( F2, 39.4%; F3, 24.1%; and F4, 8.7%) and also values seen in cohorts of patients with T2DM,<sup>44</sup> and the general population.<sup>45</sup> For a diagnosis of F2, F3, and F4, there was a marked reduction in the positive predictive value as the prevalence of fibrosis was lowered (Table S5).

## Discussion

This IPD-MA, which included data-independent cohorts from Europe, Asia, and the USA with a total of 798 patients, showed that MRE has a good or excellent diagnostic performance with AUROCs of 0.82, 0.92, 0.92, and 0.94 for staging mild ( F1), significant ( F2), or advanced ( F3) fibrosis and cirrhosis (F4), respectively, in patients with NAFLD. However, discordance between MRE and pathological findings was still observed in some patients. It is recognised that this may, at least in part, be attributable to histology being an imperfect reference standard.<sup>6,46</sup> Although the AUROC has been the most recommended and applied method for evaluating the performance of numerical diagnostic tests, in this study, we also used multilevel modelling approaches (LMM and GLMM) to analyse the confounding factors.<sup>47</sup> Final binary outcomes were evaluated based on the misclassification results, which may be more clinically relevant. Compared with the conventional linear and logistic regression models, LMM and GLMM are now more widely accepted statistical analysis tools using hierarchical (random-effects) data that account for the clustering of patients within studies.<sup>27</sup> LMM consistently identified age, GGT level, platelet count, greater steatohepatitis grade, and liver fibrosis stage to be associated with LSM by MRE. Platelet count and age had been included in several indices, such as aspartate transaminase (AST)-to-platelet ratio index, fibrosis-4 index, and Hepamet fibrosis score for the non-invasive prediction of hepatic fibrosis.<sup>7,48</sup> GGT is present in the bile canaliculi of hepatocytes and biliary epithelial cells. It is strongly associated with oxidative stress, which may contribute to clinical progression from simple fatty liver to NASH.<sup>49</sup> In addition, a recent clinical trial of paediatric NASH<sup>50</sup> showed that GGT had a stronger relationship with improvement in NAS than aminotransferase (ALT and AST). GGT was associated with the development of metabolic syndrome (MetS) in a non-linear dose-response relationship.<sup>51</sup> It also can be considered a more sensitive and stronger biomarker of MetS than aminotransferases (ALT and AST). A previous IPD-MA<sup>52</sup> with a large cohort (n = 16,802) showed that significant increases in liver stiffness by TE were associated with MetS. This may suggest that patients with NAFLD with MetS may exhibit higher liver stiffness than patients with NAFLD without MetS. The impact of inflammation on ultrasound elastography biomarkers had been reported in two previous IPD-MAs on TE<sup>53</sup> and 2D SWE<sup>54</sup> for diagnosing liver fibrosis, which used AST and ALT to indicate liver inflammatory activity rather than pathological findings. Furthermore, similar results were found in several single-centre studies on MRE,<sup>55-57</sup> where higher inflammation activity was associated with higher LSM by MRE. Tissue inflammation increases local blood availability, inflammatory cell infiltration, and interstitial pressure, which may increase liver stiffness.<sup>57</sup> The results of our study are consistent with the existing literature, which indicates that inflammatory activity can increase LSM by MRE in patients showing the same stage of fibrosis, as we demonstrated by taking into account inflammatory activity in the liver tissue. However, the main aetiology of patients included in these previous studies was viral hepatitis. Serum aminotransferases are routinely measured to detect liver disease, but their specificity and sensitivity for NASH are low.

The average liver stiffness increased among different activity NASH and GGT concentration, but only that at the F0–F1 stage was statistically significant, which may

be because of the unbalanced distribution of liver fibrosis (F0–F1 prevalence, 60.6%). The difference in stiffness between stages 0, 1, and 2 was much smaller than that between stages 2, 3, and 4. Accordingly, the effect of incremental stiffness caused by inflammation would be expected to be most prominent at lower stages of fibrosis. However, it is worth noting that a certain number of cases with high inflammatory activity were still seen in underestimated patients. In contrast, the higher inflammatory activity caused by viral hepatitis may affect all stages of liver fibrosis.<sup>55-57</sup> There was also no interaction between NASH-SA and GGT. GGT was a biomarker lacking high specificity, and its elevation could not be explained by the inflammation severity. Furthermore, we observed that the LSM of no-NASH appeared to be higher than that of NASH-MMA in advanced fibrosis, which may be because of the appearance of ‘burn-out in NASH-related cirrhosis’ and its misclassification as no-NASH.<sup>58</sup> These results also provide motivation for further exploration of additional biomarkers provided by advanced 3D vector MRE, which shows promise for the independent distinction between fibrosis and inflammation.<sup>59</sup>

In addition to the previously mentioned details, the T2DM presence was shown to potentially influence diagnostic performance in subgroup analysis and univariate analysis of GLMM. A strong association between increased LSM by TE and the presence of diabetes mellitus and/or greater insulin resistance was observed not only in the whole population but also in a subgroup of participants with NAFLD.<sup>60</sup> This may apparently induce patients with NAFLD with T2DM, which positively affects LSM in NAFLD. Although iron overload is the most dominant cause of technical failure of MRE<sup>9</sup> and could be a confounding factor, when patients with  $R2^* > 80$  Hz in sensitivity analysis were removed, MRE remained as efficient as it previously was in detecting advanced fibrosis and cirrhosis; however, the number of patients with  $R2^*$  information was limited.

Up until now, several conventional meta-analyses<sup>61,62</sup> have gathered the findings and listed different cut-off values for detecting liver fibrosis in patients with NAFLD. However, these study-level meta-analyses were unable to establish new cut-offs without access to the original data; the widely accepted cut-off value for MRE still remained an unmet need. Although the optimal cut-off values were calculated by two previous IPD-MAs,<sup>14,15</sup> these studies included only 230 and 232 patients with NAFLD, respectively.

There are several limitations to our study. First, the criteria for NASH diagnosis are subtly different between the two scoring systems used (SAF and NASH-CRN), which may result in misclassification; in addition, the classification of inflammatory activity for NASH is also an artificial harmonisation with significant subjectivity, which may introduce selection bias. Second, although IPD pooled analysis was able to alleviate several of the limitations of a conventional aggregate data meta-analysis, ours was still a retrospective analysis with several inherent variations as a result of the lack of standardised performance of index tests and the lack of centralised re-reading of biopsies. Third, liver biopsy itself is not a perfect gold standard as it samples only one in 50,000 of total liver mass and a significant discrepancy in the fibrosis stage as high as 33% can be observed depending on the site of liver biopsy.<sup>6,46</sup> Thus, potential bias regarding the interpretation of the liver biopsy cannot be excluded. Fourth, although the validation group was generated by random sampling, it was still difficult to avoid all the influence of overlapping data on the results. Fifth, as

significant heterogeneity was present, it was not explained by meta-regression and subgroup analysis. We therefore chose GLMM and LMM instead of traditional linear and logistic regression analysis. Sixth, note also that MRE values were used as another endpoint, to define covariates that affected liver stiffness, irrespective of the diagnosis.

In conclusion, through an IPD-MA, we observed that MRE is a highly accurate, non-invasive technique for staging liver fibrosis in patients with NAFLD and NASH, where we have established cut-offs of 2.65, 3.14, 3.53, and 4.45 kPa for mild ( F1), significant ( F2), and advanced ( F3) fibrosis, and cirrhosis (F4), respectively. Nonetheless, clinical information and the possible presence of severe inflammation activity should be considered for early-stage fibrosis cohorts to optimise the diagnostic accuracy of MRE in staging liver fibrosis.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgements

The authors thank all patients and clinical researchers who were involved in the studies of this meta-analysis of individual patient data, and the authors acknowledge Sara Romero-Otero for editorial assistance.

## Financial support

We would like to thank LITMUS (Liver Investigation: Testing Marker Utility in Steatohepatitis) project for supporting this research. The LITMUS study is funded by the Innovative Medicines Initiative 2 (IMI2) Joint Undertaking under grant agreement 777377. This Joint Undertaking receives support from the European Union's Horizon 2020 research and innovation programme and the European Federation of Pharmaceutical Industries and Associations (EFPIA; [efpia.edu](http://efpia.edu)). RL receives funding support from NCATS (5UL1TR001442), NIDDK (U01DK061734, U01DK130190, R01DK106419, R01DK121378, R01DK124318, and P30DK120515), NHLBI (P01HL147835), and NIAAA (U01AA029019). QMA is a coordinator of the LITMUS consortium and a Newcastle NIHR Biomedical Research Centre investigator. MP acknowledges support from the Oxford NIHR Biomedical Research Centre. RLE acknowledges support from the National Institutes of Health (grant R37 EB001981). DHL received a grant from KHIDI (Korea Health Industry Development Institute) (HI14C1135), funded by the Ministry of Health and Welfare, Korea. JB acknowledges funding support from NCATS (4UH3TR003289), NCI 1R01CA255809, and NIAAA (5U01AA026978).

## Data availability statement

Data are available on reasonable request. Interested parties should contact the corresponding author to request specific data.

## Abbreviations

|              |  |
|--------------|--|
| <b>ALT</b>   | alanine aminotransferase                               |
| <b>AST</b>   | aspartate aminotransferase                             |
| <b>AUROC</b> | area under the receiver operating characteristic curve |
| <b>GGT</b>   | gamma-glutamyl transferase                             |
| <b>GLMM</b>  | generalised linear mixed model                         |

|               |                                       |
|---------------|---------------------------------------|
| <b>GRE</b>    | gradient-echo sequence                |
| <b>IPD</b>    | individual patient data               |
| <b>IPD-MA</b> | individual patient data meta-analysis |
| <b>LMM</b>    | linear mixed-effects model            |
| <b>LSM</b>    | liver stiffness measurement           |
| <b>MetS</b>   | metabolic syndrome                    |
| <b>MR</b>     | magnetic resonance                    |
| <b>MRE</b>    | magnetic resonance elastography       |
| <b>MRI</b>    | magnetic resonance imaging            |
| <b>NAFLD</b>  | non-alcoholic fatty liver disease     |
| <b>NAS</b>    | NAFLD activity score                  |
| <b>MMA</b>    | mild–moderate activity                |
| <b>SA</b>     | severe activity                       |
| <b>OR</b>     | odds ratio                            |
| <b>SE-EPI</b> | spin-echo echo-planar imaging         |
| <b>SAF</b>    | steatosis–activity–fibrosis           |
| <b>SWE</b>    | shear wave elastography               |
| <b>TE</b>     | transient elastography                |
| <b>T2DM</b>   | type 2 diabetes mellitus              |

## References

Author names in bold designate shared co-first authorship

- [1]. Younossi ZM, Koenig AB, Abdelatif D, Fazel Y, Henry L, Wymer M. Global epidemiology of nonalcoholic fatty liver disease – meta-analytic assessment of prevalence, incidence, and outcomes. *Hepatology* 2016;64:73–84. [PubMed: 26707365]
- [2]. Wong T, Wong RJ, Gish RG. Diagnostic and treatment implications of nonalcoholic fatty liver disease and nonalcoholic steatohepatitis. *Gastroenterol Hepatol* 2019;15:83–89.
- [3]. Taylor RS, Taylor RJ, Bayliss S, Hagström H, Nasr P, Schattenberg JM, et al. Association between fibrosis stage and outcomes of patients with nonalcoholic fatty liver disease: a systematic review and meta-analysis. *Gastroenterology* 2020;158:1611–1625.e12. [PubMed: 32027911]
- [4]. Sanyal AJ. Past, present and future perspectives in nonalcoholic fatty liver disease. *Nat Rev Gastroenterol Hepatol* 2019;16:377–386. [PubMed: 31024089]
- [5]. Takyar V, Etzion O, Heller T, Kleiner DE, Rotman Y, Ghany MG, et al. Complications of percutaneous liver biopsy with Klatskin needles: a 36-year single-centre experience. *Aliment Pharmacol Ther* 2017;45:744–753. [PubMed: 28074540]

- [6]. Brunt EM, Clouston AD, Goodman Z, Guy C, Kleiner DE, Lackner C, et al. Complexity of ballooned hepatocyte feature recognition: defining a training atlas for artificial intelligence-based imaging in NAFLD. *J Hepatol* 2022;76:1030–1041. [PubMed: 35090960]
- [7]. Ampuero J, Pais R, Aller R, Gallego-Durán R, Crespo J, García-Monzón C, et al. Development and validation of Hepamet fibrosis scoring system – a simple, noninvasive test to identify patients with nonalcoholic fatty liver disease with advanced fibrosis. *Clin Gastroenterol Hepatol* 2020;18:216–225.e5. [PubMed: 31195161]
- [8]. Papatheodoridi M, Cholongitas E. Diagnosis of non-alcoholic fatty liver disease (NAFLD): current concepts. *Curr Pharm Des* 2018;24:4574–4586. [PubMed: 30652642]
- [9]. Hoodeshenas S, Yin M, Venkatesh SK. Magnetic resonance elastography of liver – current update. *Top Magn Reson Imaging* 2018;27:319–333. [PubMed: 30289828]
- [10]. Imajo K, Kessoku T, Honda Y, Tomeno W, Ogawa Y, Mawatari H, et al. Magnetic resonance imaging more accurately classifies steatosis and fibrosis in patients with nonalcoholic fatty liver disease than transient elastography. *Gastroenterology* 2016;150:626–637.e7. [PubMed: 26677985]
- [11]. Park CC, Nguyen P, Hernandez C, Bettencourt R, Ramirez K, Fortney L, et al. Magnetic resonance elastography vs transient elastography in detection of fibrosis and noninvasive measurement of steatosis in patients with biopsy-proven nonalcoholic fatty liver disease. *Gastroenterology* 2017;152:598–607.e2. [PubMed: 27911262]
- [12]. Caussy C, Chen J, Alquiraish MH, Cepin S, Nguyen P, Hernandez C, et al. Association between obesity and discordance in fibrosis stage determination by magnetic resonance vs transient elastography in patients with nonalcoholic liver disease. *Clin Gastroenterol Hepatol* 2018;16:1974–1982.e7.
- [13]. Zhang YN, Fowler KJ, Boehringer AS, Montes V, Schlein AN, Covarrubias Y, et al. Comparative diagnostic performance of ultrasound shear wave elastography and magnetic resonance elastography for classifying fibrosis stage in adults with biopsy-proven nonalcoholic fatty liver disease. *Eur Radiol* 2022;32:2457–2469. [PubMed: 34854929]
- [14]. Singh S, Venkatesh SK, Loomba R, Wang Z, Sirlin C, Chen J, et al. Magnetic resonance elastography for staging liver fibrosis in non-alcoholic fatty liver disease: a diagnostic accuracy systematic review and individual participant data pooled analysis. *Eur Radiol* 2016;26:1431–1440. [PubMed: 26314479]
- [15]. Hsu C, Caussy C, Imajo K, Chen J, Singh S, Kaulback K, et al. Magnetic resonance vs transient elastography analysis of patients with nonalcoholic fatty liver disease: a systematic review and pooled analysis of individual participants. *Clin Gastroenterol Hepatol* 2019;17:630–637.e8. [PubMed: 29908362]
- [16]. Stewart LA, Clarke M, Rovers M, Riley RD, Simmonds M, Stewart G, et al. Preferred reporting items for a systematic review and meta-analysis of individual participant data: the PRISMA-IPD statement. *JAMA* 2015;313:1657–1665. [PubMed: 25919529]
- [17]. McInnes MD, Moher D, Thoms BD, McGrath TA, Bossuyt PM, Clifford T, et al. Preferred reporting items for a systematic review and meta-analysis of diagnostic test accuracy studies: the PRISMA-DTA statement. *JAMA* 2018;319:388–396. [PubMed: 29362800]
- [18]. 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;94:2467–2474. [PubMed: 10484010]
- [19]. Brunt EM. Nonalcoholic steatohepatitis: definition and pathology. *Semin Liver Dis* 2001;21:3–16. [PubMed: 11296695]
- [20]. Kleiner DE, Brunt EM, Van Natta M, Behling C, Contos MJ, Cummings OW, et al. Design and validation of a histological scoring system for nonalcoholic fatty liver disease. *Hepatology* 2005;41:1313–1321. [PubMed: 15915461]
- [21]. Bedossa P, Poitou C, Veyrie N, Bouillot JL, Basdevant A, Paradis V, et al. Histopathological algorithm and scoring system for evaluation of liver lesions in morbidly obese patients. *Hepatology* 2012;56:1751–1759. [PubMed: 22707395]

- [22]. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 2011;155:529–536. [PubMed: 22007046]
- [23]. Fluss R, Faraggi D, Reiser B. Estimation of the Youden Index and its associated cutoff point. *Biom J* 2005;47:458–472. [PubMed: 16161804]
- [24]. Riley RD, Dodd SR, Craig JV, Thompson JR, Williamson PR. Meta-analysis of diagnostic test studies using individual patient data and aggregate data. *Stat Med* 2008;27:6111–6136. [PubMed: 18816508]
- [25]. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ* 2003;327:557–560. [PubMed: 12958120]
- [26]. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 1988;44:837–845. [PubMed: 3203132]
- [27]. Debray TP, Moons KG, van Valkenhoef G, Efthimiou O, Hummel N, Groenwold RH, et al. Get real in individual participant data (IPD) meta-analysis: a review of the methodology. *Res Synth Methods* 2015;6:293–309. [PubMed: 26287812]
- [28]. Chen J, Talwalkar JA, Yin M, Glaser KJ, Sanderson SO, Ehman RL. Early detection of nonalcoholic steatohepatitis in patients with nonalcoholic fatty liver disease by using MR elastography. *Radiology* 2011;259:749–756. [PubMed: 21460032]
- [29]. Bours MJ. Bayes' rule in diagnosis. *J Clin Epidemiol* 2021;131:158–160. [PubMed: 33741123]
- [30]. Costa-Silva L, Ferolla SM, Lima AS, Vidigal PVT, de Abreu Ferrari TC. MR elastography is effective for the non-invasive evaluation of fibrosis and necroinflammatory activity in patients with nonalcoholic fatty liver disease. *Eur J Radiol* 2018;98:82–89. [PubMed: 29279175]
- [31]. Kim JW, Lee Y-S, Park YS, Kim B-H, Lee SY, Yeon JE, et al. Multiparametric MR index for the diagnosis of non-alcoholic steatohepatitis in patients with non-alcoholic fatty liver disease. *Sci Rep* 2020;10:2671. [PubMed: 32060386]
- [32]. Lee Y-S, Yoo YJ, Jung YK, Kim JH, Seo YS, Yim HJ, et al. Multiparametric MR is a valuable modality for evaluating disease severity of nonalcoholic fatty liver disease. *Clin Transl Gastroenterol* 2020;11:e00157. [PubMed: 32251018]
- [33]. Furlan A, Tublin ME, Yu L, Chopra KB, Lippello A, Behari J. Comparison of 2D shear wave elastography, transient elastography, and MR elastography for the diagnosis of fibrosis in patients with nonalcoholic fatty liver disease. *AJR Am J Roentgenol* 2020;214:W20–W26. [PubMed: 31714842]
- [34]. Han MAT, Vipani A, Noureddin N, Ramirez K, Gornbein J, Saouaf R, et al. MR elastography-based liver fibrosis correlates with liver events in nonalcoholic fatty liver patients: a multicenter study. *Liver Intern* 2020;40:2242–2251.
- [35]. Choi SJ, Kim SM, Kim YS, Kwon OS, Shin SK, Kim KK, et al. Magnetic resonance-based assessments better capture pathophysiologic profiles and progression in nonalcoholic fatty liver disease. *Diabetes Metab J* 2021;45:739–752. [PubMed: 33108854]
- [36]. Cui J, Ang B, Haufe W, Hernandez C, Verna EC, Sirlin CB, et al. Comparative diagnostic accuracy of magnetic resonance elastography vs. eight clinical prediction rules for non-invasive diagnosis of advanced fibrosis in biopsy-proven non-alcoholic fatty liver disease: a prospective study. *Aliment Pharmacol Ther* 2015;41:1271–1280. [PubMed: 25873207]
- [37]. Loomba R, Wolfson T, Ang B, Hooker J, Behling C, Peterson M, et al. Magnetic resonance elastography predicts advanced fibrosis in patients with nonalcoholic fatty liver disease: a prospective study. *Hepatology* 2014;60:1920–1928. [PubMed: 25103310]
- [38]. Cui J, Heba E, Hernandez C, Haufe W, Hooker J, Andre MP, et al. Magnetic resonance elastography is superior to acoustic radiation force impulse for the diagnosis of fibrosis in patients with biopsy-proven nonalcoholic fatty liver disease: a prospective study. *Hepatology* 2016;63:453–461. [PubMed: 26560734]
- [39]. Loomba R, Cui J, Wolfson T, Haufe W, Hooker J, Szeverenyi N, et al. Novel 3D magnetic resonance elastography for the noninvasive diagnosis of advanced fibrosis in NAFLD: a prospective study. *Am J Gastroenterol* 2016;111:986–994. [PubMed: 27002798]



- [40]. Loomba R, Lam J, Wolfson T, Ang B, Bhatt A, Peterson MR, et al. Magnetic resonance elastography accurately predicts advanced fibrosis in NAFLD: a pilot study with paired-liver biopsy. *Gastroenterology* 2013;144:S1012.
- [41]. Kim D, Kim WR, Talwalkar JA, Kim HJ, Ehman RL. Advanced fibrosis in nonalcoholic fatty liver disease: noninvasive assessment with MR elastography. *Radiology* 2013;268:411–419. [PubMed: 23564711]
- [42]. Lara Romero C, Liang JX, Fernández Lizaranzazu I, Ampuero Herrojo J, Castell J, Del Prado Alba C, et al. Liver stiffness accuracy by MR elastography in histologically proven non-alcoholic fatty liver disease patients: a Spanish cohort. *Rev Esp Enferm Dig* 2023;115:162–167. [PubMed: 35791792]
- [43]. Alsaqal S, Hockings P, Ahlström H, Gummesson A, Hedström A, Hulthe J, et al. The combination of MR elastography and proton density fat fraction improves diagnosis of nonalcoholic steatohepatitis. *J Magn Reson Imaging* 2022;56:368–379. [PubMed: 34953171]
- [44]. Ajmera V, Cepin S, Tesfai K, Hofflich H, Cadman K, Lopez S, et al. A prospective study on the prevalence of NAFLD, advanced fibrosis, cirrhosis and hepatocellular carcinoma in people with type 2 diabetes. *J Hepatol* 2023;78:471–478. [PubMed: 36410554]
- [45]. Harris R, Harman DJ, Card TR, Aithal GP, Guha IN. Prevalence of clinically significant liver disease within the general population, as defined by non-invasive markers of liver fibrosis: a systematic review. *Lancet Gastroenterol Hepatol* 2017;2:288–297. [PubMed: 28404158]
- [46]. Davison BA, Harrison SA, Cotter G, Alkhoury N, Sanyal A, Edwards C, et al. Suboptimal reliability of liver biopsy evaluation has implications for randomized clinical trials. *J Hepatol* 2020;73:1322–1332. [PubMed: 32610115]
- [47]. Xiao G, Zhu S, Xiao X, Yan L, Yang J, Wu G. Comparison of laboratory tests, ultrasound, or magnetic resonance elastography to detect fibrosis in patients with nonalcoholic fatty liver disease: a meta-analysis. *Hepatology* 2017;66:1486–1501. [PubMed: 28586172]
- [48]. Lee J, Vali Y, Boursier J, Spijker R, Anstee QM, Bossuyt PM, et al. Prognostic accuracy of FIB-4, NAFLD fibrosis score and APRI for NAFLD-related events: a systematic review. *Liver Int* 2021;41:261–270. [PubMed: 32946642]
- [49]. Irie M, Sohda T, Iwata K, Kunimoto H, Fukunaga A, Kuno S, et al. Levels of the oxidative stress marker  $\gamma$ -glutamyltranspeptidase at different stages of nonalcoholic fatty liver disease. *J Int Med Res* 2012;40:924–933. [PubMed: 22906265]
- [50]. Newton KP, Lavine JE, Wilson L, Behling C, Vos MB, Molleston JP, et al. Alanine aminotransferase and gamma-glutamyl transpeptidase predict histologic improvement in pediatric nonalcoholic steatohepatitis. *Hepatology* 2021;73:937–951. [PubMed: 32416645]
- [51]. Kunutsor S, Apekey T, Seddoh D. Gamma glutamyltransferase and metabolic syndrome risk: a systematic review and dose–response meta-analysis. *Int J Clin Pract* 2015;69:136–144. [PubMed: 25363194]
- [52]. Bazerbachi F, Haffar S, Wang Z, Cabezas J, Arias-Loste MT, Crespo J, et al. Range of normal liver stiffness and factors associated with increased stiffness measurements in apparently healthy individuals. *Clin Gastroenterol Hepatol* 2019;17:54–64.e1.
- [53]. Nguyen-Khac E, Thiele M, Voican C, Nahon P, Moreno C, Boursier J, et al. Non-invasive diagnosis of liver fibrosis in patients with alcohol-related liver disease by transient elastography: an individual patient data meta-analysis. *Lancet Gastroenterol Hepatol* 2018;3:614–625. [PubMed: 29983372]
- [54]. Herrmann E, de Ledinghen V, Cassinotto C, Chu WC, Leung VY, Ferraioli G, et al. Assessment of biopsy-proven liver fibrosis by two-dimensional shear wave elastography: an individual patient data-based meta-analysis. *Hepatology* 2018;67:260–272. [PubMed: 28370257]
- [55]. Shi Y, Guo Q, Xia F, Dzyubak B, Glaser KJ, Li Q, et al. MR elastography for the assessment of hepatic fibrosis in patients with chronic hepatitis B infection: does histologic necroinflammation influence the measurement of hepatic stiffness? *Radiology* 2014;273:88–98. [PubMed: 24893048]
- [56]. Ichikawa S, Motosugi U, Nakazawa T, Morisaka H, Sano K, Ichikawa T, et al. Hepatitis activity should be considered a confounder of liver stiffness measured with MR elastography. *J Magn Reson Imaging* 2015;41:1203–1208. [PubMed: 24889753]

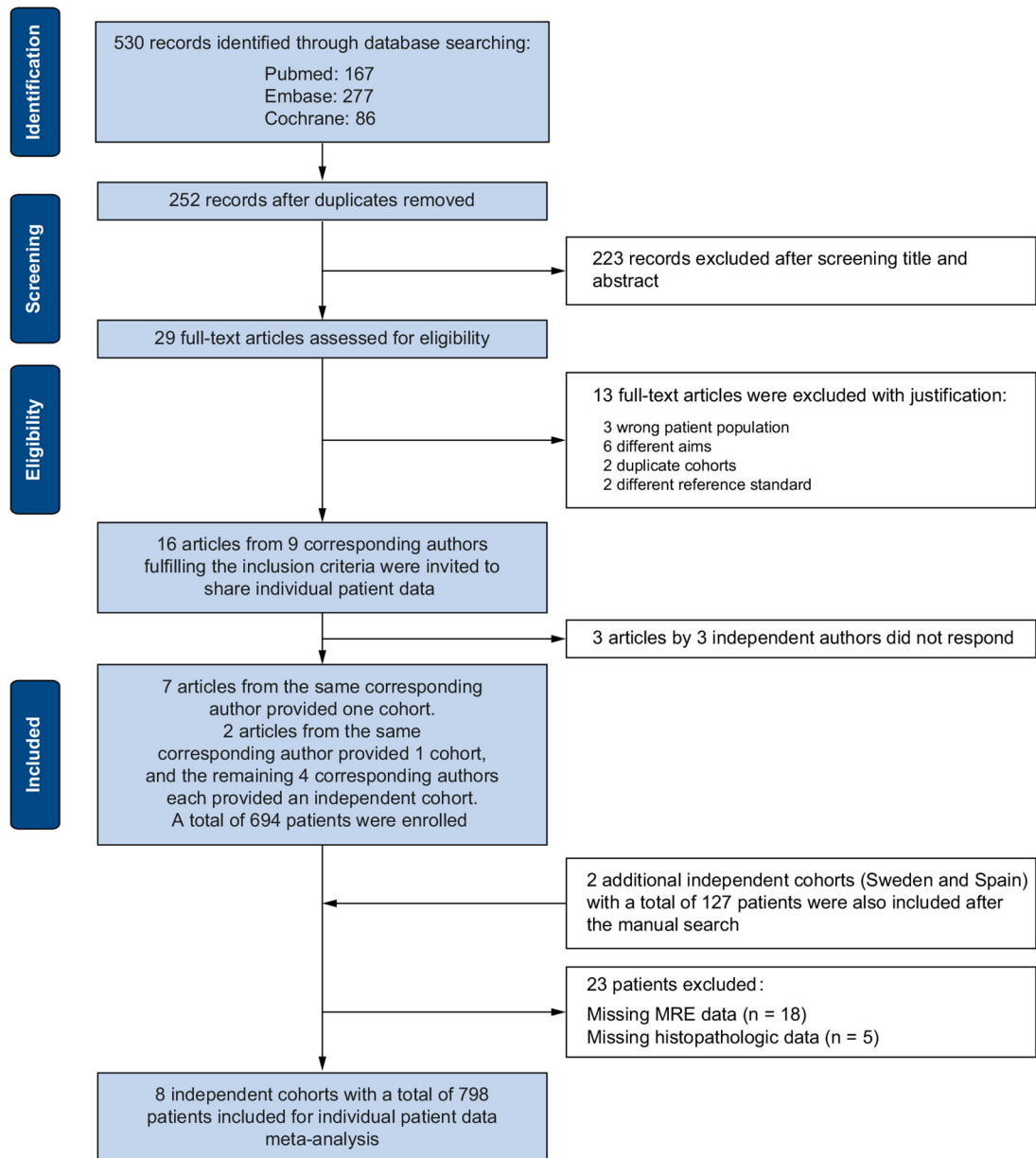
- [57]. Yin M, Glaser KJ, Talwalkar JA, Chen J, Manduca A, Ehman RL. Hepatic MR elastography: clinical performance in a series of 1377 consecutive examinations. *Radiology* 2016;278:114–124. [PubMed: 26162026]
- [58]. Ampuero J, Aller R, Gallego-Durán R, Crespo J, Abad J, González-Rodríguez Á, et al. Definite and indeterminate nonalcoholic steatohepatitis share similar clinical features and prognosis: a longitudinal study of 1893 biopsy-proven nonalcoholic fatty liver disease subjects. *Liver Intern* 2021;41:2076–2086.
- [59]. Reeder SB. Emergence of 3D MR elastography–based quantitative markers for diffuse liver disease. *Radiology* 2021;301:163–165. [PubMed: 34374596]
- [60]. Koehler EM, Plompen EP, Schouten JN, Hansen BE, Darwish Murad S, Taimr P, et al. Presence of diabetes mellitus and steatosis is associated with liver stiffness in a general population: the Rotterdam study. *Hepatology* 2016;63:138–147. [PubMed: 26171685]
- [61]. Selvaraj EA, Mózes FE, Jayaswal ANA, Zafarmand MH, Vali Y, Lee JA, et al. Diagnostic accuracy of elastography and magnetic resonance imaging in patients with NAFLD: a systematic review and meta-analysis. *J Hepatol* 2021;75:770–785. [PubMed: 33991635]
- [62]. Li J, Fine JP, Pencina MJ. Multi-category diagnostic accuracy based on logistic regression. *Stat Theor Relat Fields* 2017;1:143–158.

### Highlights

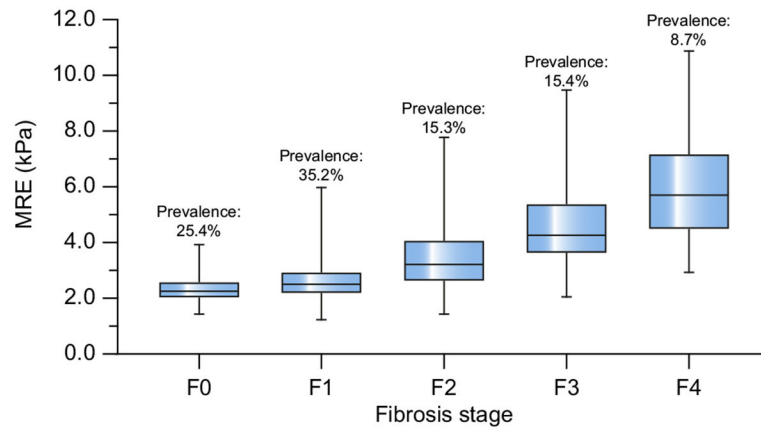
- Magnetic resonance elastography demonstrated excellent diagnostic accuracy for staging liver fibrosis in patients with NAFLD.
- Assessment of fibrosis was not confounded by steatosis or high body mass index.
- Cut-offs for significant fibrosis, advanced fibrosis, and cirrhosis were found to be 3.14, 3.53, and 4.45 kPa, respectively.
- Raised GGT levels and increased lobular inflammatory activity may result in overestimation of early-stage fibrosis (F0–F1).

### Impact and implications

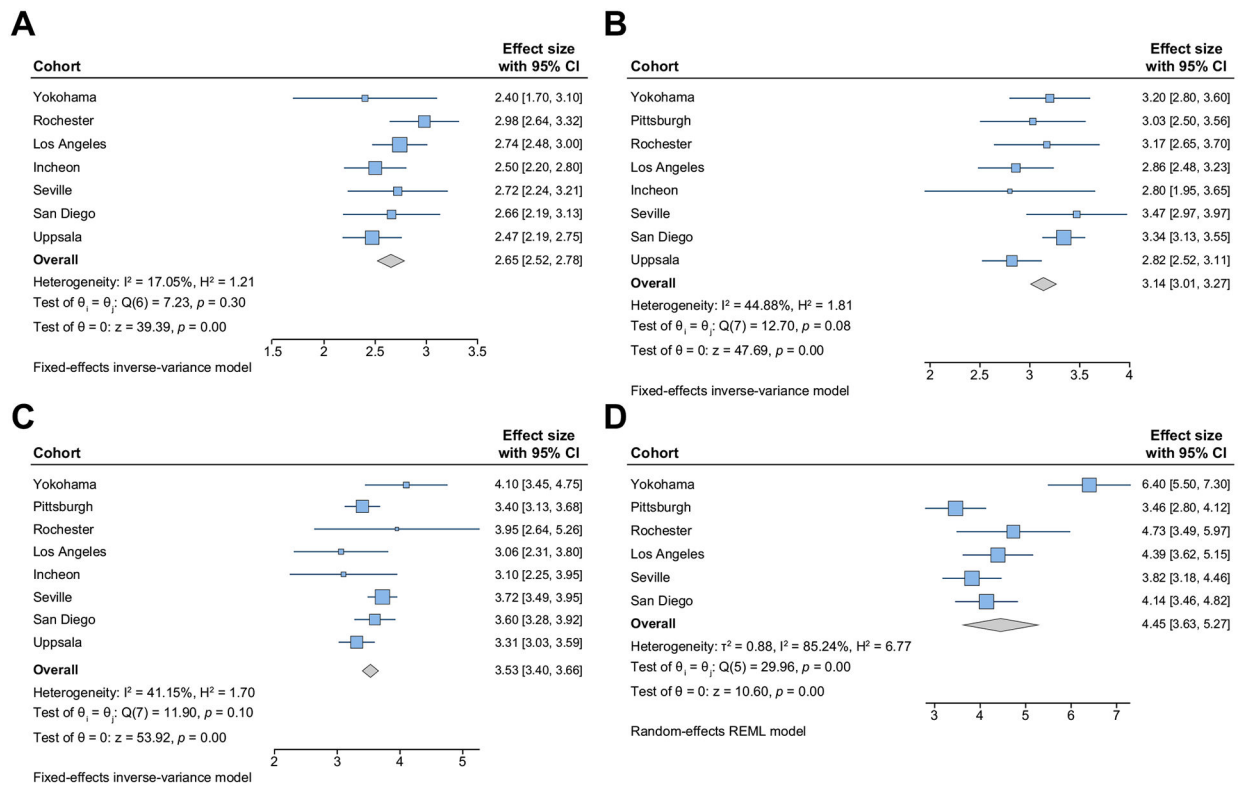
This individual patient data meta-analysis of eight international cohorts, including 798 patients, demonstrated that MRE achieves excellent diagnostic accuracy for significant, advanced fibrosis and cirrhosis in patients with NAFLD. Cut-off values (significant fibrosis, 3.14 kPa; advanced fibrosis, 3.53 kPa; and cirrhosis, 4.45 kPa) were established. Elevated inflammatory activity and gamma-glutamyltransferase level may affect the diagnostic accuracy of MRE, leading to overestimation of liver fibrosis in early stages. We observed no impact of diabetes, obesity, or any other metabolic disorder on the diagnostic accuracy of MRE.



**Fig. 1. Study identification and selection flow chart.**  
MRE, magnetic resonance elastography.



**Fig. 2. Liver stiffness values by MRE corresponding to each liver fibrosis stage.** The bars represent the minimum and maximum; the top and bottom lines in the box represent the 25% and 75% levels, respectively; and the middle line indicates the median level. MRE, magnetic resonance elastography.



**Fig. 3. Pooled cut-offs for staging liver fibrosis stage in patients with NAFLD.** (A) Pooled cut-off for mild fibrosis ( F1). (B) Pooled cut-off for significant fibrosis ( F2). (C) Pooled cut-off for advanced fibrosis ( F3). (D) Pooled cut-off for cirrhosis (F4). Pooled cut-offs were estimated by study-level cut-offs in fixed or random-effects model meta-analyses and were displayed graphically as forest plots. The diamond symbol represents the pooled cut-off value, and its width represents the 95% CI. REML, restricted maximum likelihood.

**Table 1.**

Study characteristics of each independent cohort.

| Centre                           | Country | Study design  | Duration                       | Patients (n) | Field strengths (T) | Manufacturer | Sequence | Histological criteria | Interval between MRE and biopsy (days), median (IQR) |
|----------------------------------|---------|---------------|--------------------------------|--------------|---------------------|--------------|----------|-----------------------|--|
| Rochester <sup>28</sup>          | USA     | Retrospective | January 2007 to March 2010     | 58           | 1.5                 | GE           | GRE      | Brunt                 | 31 (9–40)  |
| Yokohama <sup>10</sup>           | Japan   | Prospective   | July 2013 to April 2015        | 142          | 3.0                 | GE           | SE-EPI   | NASH-CRN              | 60 (24–121)  |
| Incheon <sup>35</sup>            | Korea   | Prospective   | June 2017 to May 2020          | 82           | 3.0                 | Siemens      | SE-EPI   | NASH-CRN              | 1 (1–7)  |
| Pittsburgh <sup>33</sup>         | USA     | Prospective   | October 2015 to December 2017  | 59           | 1.5                 | GE           | GRE      | NASH-CRN              | 82 (46–150)  |
| Los Angeles <sup>34</sup>        | USA     | Retrospective | May 2016 to March 2019         | 101          | 3.0                 | Siemens      | SE-EPI   | NASH-CRN              | 92 (39–158)  |
| Seville <sup>42</sup>            | Spain   | Prospective   | November 2019 to December 2020 | 61           | 3.0                 | Philips      | GRE      | SAF                   | 77 (38–119)  |
| San Diego <sup>11,12,36-40</sup> | USA     | Prospective   | January 2012 to June 2020      | 233          | 3.0                 | GE           | GRE      | NASH-CRN              | 34 (14–71)   |
| Uppsala <sup>43</sup>            | Sweden  | Prospective   | March 2017 to December 2019    | 62           | 3.0                 | GE           | SE-EPI   | SAF                   | 56.5 (37–79)   |

GRE, gradient-echo sequence; CRN, Clinical Research Network; SAF, steatosis-activity-fibrosis; SE-EPI, spin-echo echo-planar imaging.



**Table 2.**

Demographic details of the entire cohort.

|                                | All (n = 798)                             | F0 (n = 203, 25.4%)             | F1 (n = 281, 35.2%)             | F2 (n = 122, 15.3%)           | F3 (n = 123, 15.4%)           | F4 (n = 69, 8.7%)         |
|--------------------------------|---|---------------------------------|---------------------------------|-------------------------------|-------------------------------|---------------------------|
| Age (years)                    | 51.62 ± 14.77                             | 47.25 ± 13.86                   | 48.05 ± 15.05                   | 53.04 ± 14.69                 | 60.68 ± 11.67                 | 60.31 ± 9.35              |
| Sex (male/%)                   | 353 (44)                                  | 90 (44)                         | 126 (45)                        | 63 (52)                       | 45 (37)                       | 29 (42)                   |
| Weight (kg)                    | 89.05 ± 20.86                             | 86.91 ± 17.88                   | 91.21 ± 21.13                   | 91.66 ± 22.38                 | 85.63 ± 21.82                 | 87.26 ± 21.48             |
| BMI (n = 764)                  | 31.89 ± 5.97                              | 30.84 ± 5.42                    | 32.56 ± 6.05                    | 31.92 ± 5.82                  | 31.80 ± 6.37                  | 32.14 ± 6.38              |
| <30 kg/m <sup>2</sup>          | 304 (40)                                  | 77 (42)                         | 103 (38)                        | 49 (40)                       | 50 (41)                       | 25 (38)                   |
| 30–35 kg/m <sup>2</sup>        | 249 (33)                                  | 69 (37)                         | 79 (29)                         | 44 (36)                       | 35 (29)                       | 22 (33)                   |
| >35 kg/m <sup>2</sup>          | 211 (27)                                  | 38 (21)                         | 89 (33)                         | 29 (24)                       | 36 (30)                       | 19 (29)                   |
| T2DM (Y/%; n = 738)            | 310 (42)                                  | 34 (20)                         | 103 (39)                        | 55 (45)                       | 76 (64)                       | 42 (67)                   |
| AHT (Y/%; n = 504)             | 196 (39)                                  | 20 (27)                         | 57 (29)                         | 41 (45)                       | 53 (56)                       | 25 (52)                   |
| ALT (U/L)                      | 46.0 (30.0–73.0)                          | 37.0 (27.5–58.0)                | 47.0 (30.0–79.0)                | 53.0 (33.0–79.0)              | 54.5 (37.0–79.0)              | 38.0 (28.0–53.0)          |
| AST (U/L)                      | 38.2 (26.0–57.0)                          | 29.0 (20.0–36.0)                | 40.0 (27.5–60.5)                | 45.0 (30.0–69.4)              | 50.0 (30.7–69.00)             | 40.0 (26.0–55.0)          |
| AST/ALT                        | 0.82 (0.62–1.09)                          | 0.74 (0.59–0.97)                | 0.80 (0.57–1.13)                | 0.81 (0.63–1.09)              | 0.91 (0.71–1.12)              | 1.00 (0.74–1.23)          |
| GGT (U/L)                      | 55.0 (33.0–98.3)                          | 38.5 (23.5–83.0)                | 52.5 (30.9–87.1)                | 50.5 (31.0–88.8)              | 64.0 (39.0–127.5)             | 96.0 (58.5–158.8)         |
| Albumin (g/dl)                 | 4.42 ± 0.51                               | 4.50 ± 0.39                     | 4.48 ± 0.39                     | 4.42 ± 0.74                   | 4.34 ± 0.42                   | 4.07 ± 0.71               |
| Glucose (mg/dl)                | 111.69 ± 44.38                            | 103.35 ± 33.05                  | 110.49 ± 41.77                  | 117.61 ± 51.26                | 117.58 ± 53.44                | 117.62 ± 47.08            |
| Triglycerides (mg/dl)          | 163.64 ± 94.69                            | 153.83 ± 89.05                  | 169.75 ± 97.88                  | 162.68 ± 80.25                | 182.74 ± 117.96               | 138.04 ± 69.02            |
| Cholesterol (mg/dl)            | 183.72 ± 41.76                            | 186.67 ± 37.63                  | 189.96 ± 42.72                  | 181.52 ± 41.18                | 176.53 ± 43.37                | 166.86 ± 40.85            |
| Platelet (×10 <sup>9</sup> /L) | 242.45 ± 79.49                            | 259.09 ± 75.74                  | 261.68 ± 79.18                  | 234.39 ± 67.95                | 207.02 ± 74.46                | 197.95 ± 79.67            |
| HDL cholesterol (mg/dl)        | 46.01 ± 17.67                             | 46.75 ± 12.89                   | 46.10 ± 19.71                   | 43.36 ± 13.11                 | 46.50 ± 14.89                 | 47.70 ± 28.89             |
| LDL cholesterol (mg/dl)        | 103.20 ± 42.22                            | 107.61 ± 36.53                  | 110.57 ± 44.67                  | 100.30 ± 38.83                | 90.36 ± 44.31                 | 89.43 ± 40.10             |
| MRE (kPa)                      | 3.26 ± 1.51                               | 2.32 ± 0.47                     | 2.63 ± 0.71                     | 3.47 ± 1.16                   | 4.58 ± 1.39                   | 5.91 ± 1.84               |
| MRI-PDFF (%)                   | 14.62 ± 8.74                              | 12.66 ± 8.36                    | 17.16 ± 9.38                    | 15.94 ± 8.33                  | 13.24 ± 6.81                  | 10.16 ± 7.33              |
| R2* (Hz; n = 257)              | 55.0 (48.2–64.7)                          | 51.0 (46.5–55.5)                | 59.3 (51.5–67.3)                | 57.2 (51.8–69.0)              | 55.2 (48.3–58.7)              | 47.8 (42.6–53.3)          |
| Steatosis (0/1/2/3)            | 44 (5.5)/341 (42.7)/299 (37.5)/114 (14.3) | 35 (17)/97 (48)/49 (24)/22 (11) | 5 (2)/106 (38)/128 (45)/42 (15) | 1 (1)/47 (39)/49 (40)/25 (20) | 0 (0)/56 (45)/49 (40)/18 (15) | 3 (4)/35(51)/24(33)/7(12) |
| NASH (n = 740; no/MMA/SA)      | 315 (39)/172 (27)/253 (34)                | 124 (71)/29 (17)/21 (12)        | 99 (37)/79 (30)/88 (33)         | 29 (24)/40 (34)/50 (42)       | 39 (33)/20 (17)/59 (50)       | 24(39)/4(6)/35(55)        |

Continuous variables are displayed as mean ± SD, or median (IQR). Categorical variables are displayed as n (%). ‘(%)’ represents the percentage in this fibrosis stage.

AHT, arterial hypertension; ALT, alanine aminotransferase, AST, aspartate aminotransferase, GGT, gamma-glutamyltransferase, MMA, mild-moderate activity; MRE, magnetic resonance elastography; MRI-PDFF, magnetic resonance imaging proton density fat fraction; NASH, non-alcoholic steatohepatitis; SA, severe activity; T2DM, type 2 diabetes mellitus.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**Table 3.**

The diagnostic performance in each independent cohort and in meta-analysis.

|              | Pittsburgh<br>(n = 59) | Incheon<br>(n = 82) | Yokohama<br>(n = 142) | Rochester<br>(n = 58) | Seville<br>(n = 61) | Los Angeles<br>(n = 101) | San Diego<br>(n = 233) | Uppsala<br>(n = 62) | Meta-analysis<br>(95% CI) | I <sup>2</sup> |
|--------------|------------------------|---------------------|-----------------------|-----------------------|---------------------|--------------------------|------------------------|---------------------|---------------------------|----------------|
| <b>F1</b>    |                        |                     |                       |                       |                     |                          |                        |                     |                           |                |
| Prevalence   | 98.0%                  | 70.7%               | 90.1%                 | 50%                   | 78.7%               | 81.2%                    | 57.0%                  | 93.5%               | 74.6%                     |                |
| Cut-off, kPa | ND                     | 2.50 (2.20–2.80)    | 2.40 (2.20–3.60)      | 2.98 (2.5–3.17)       | 2.72 (2.66–3.63)    | 2.74 (2.7–3.23)          | 2.66 (2.28–3.22)       | 2.47 (2.25–2.82)    | 2.65 (2.52–2.78)          | 17.05%         |
| AUROC        | ND                     | 0.65 (0.53–0.77)    | 0.85 (0.78–0.91)      | 0.88 (0.79–0.96)      | 0.87 (0.76–0.97)    | 0.84 (0.75–0.91)         | 0.81 (0.75–0.86)       | 0.58 (0.34–0.81)    | 0.82 (0.79–0.85)          |                |
| Sensitivity  | ND                     | 47% (33–60%)        | 77% (68–84%)          | 83% (64–94%)          | 85% (72–94%)        | 71% (60–80%)             | 61% (52–70%)           | 57% (43–70%)        | 69% (59–78%)              | 81.84%         |
| Specificity  | ND                     | 88% (68–97%)        | 86% (64–100%)         | 77% (58–90%)          | 85% (55–98%)        | 95% (74–100%)            | 80% (69–86%)           | 75% (19–99%)        | 82% (76–87%)              | 0.00%          |
| PLR          | ND                     | 3.72 (1.2–11.1)     | 5.46 (1.51–19.75)     | 4.00 (1.9–8.3)        | 5.55 (1.5–20.0)     | 13.44 (2.0–91.0)         | 3.86 (2.4–6.2)         | 2.28 (0.4–12.6)     | 3.8 (2.70–5.30)           |                |
| NLR          | ND                     | 0.61 (0.5–0.8)      | 0.26 (0.17–0.38)      | 0.22 (0.10–0.5)       | 0.17 (0.08–0.4)     | 0.31 (0.2–0.4)           | 0.46 (0.4–0.6)         | 0.57 (0.3–1.1)      | 0.38 (0.27–0.52)          |                |
| DOR          | ND                     | 6.1 (1.6–22.7)      | 21.49 (4.5–101.4)     | 18.4 (4.9–68.7)       | 32.2 (5.8–177.5)    | 43.5 (5.5–344.4)         | 7.6 (4.1–14.2)         | 4.0 (0.4–40.4)      | 10.0 (6.0–18.0)           |                |
| <b>F2</b>    |                        |                     |                       |                       |                     |                          |                        |                     |                           |                |
| Prevalence   | 71.2%                  | 15.9%               | 54.2%                 | 24.1%                 | 50.8%               | 42.6%                    | 27.7%                  | 43.5%               | 39.4%                     |                |
| Cut-off, kPa | 3.03 (2.35–3.4)        | 2.80 (2.3–4.0)      | 3.20 (2.80–3.60)      | 3.17 (2.90–3.95)      | 3.47 (2.72–3.72)    | 2.86 (2.5–3.25)          | 3.34 (3.31–3.73)       | 2.82 (2.40–2.99)    | 3.14 (3.01–3.24)          | 44.88%         |
| AUROC        | 0.85 (0.74–0.95)       | 0.74 (0.57–0.91)    | 0.92 (0.86–0.96)      | 0.94 (0.84–0.98)      | 0.88 (0.79–0.97)    | 0.91 (0.83–0.96)         | 0.93 (0.89–0.97)       | 0.74 (0.61–0.87)    | 0.92 (0.90–0.94)          |                |
| Sensitivity  | 69% (55–81%)           | 54% (25–81%)        | 86% (78–94%)          | 100% (77–100%)        | 81% (63–93%)        | 93% (81–98%)             | 76% (63–85%)           | 52% (32–71%)        | 79% (67–88%)              | 75.00%         |
| Specificity  | 88% (64–98%)           | 91% (82–97%)        | 87% (78–95%)          | 73% (57–85%)          | 83% (65–94%)        | 79% (67–89%)             | 95% (91–98%)           | 91% (77–98%)        | 89% (82–94%)              | 67.64%         |
| PLR          | 4.94 (1.54–15.88)      | 6.20 (2.5–15.5)     | 6.72 (3.57–12.64)     | 3.67 (2.3–5.9)        | 4.84 (2.1–11.0)     | 4.50 (2.7–7.5)           | 26.15 (10.9–62.7)      | 6.05 (1.9–18.9)     | 7.30 (4.90–10.80)         |                |
| NLR          | 0.37 (0.23–0.59)       | 0.51 (0.3–0.9)      | 0.14 (0.077–0.256)    | 0.05 (0.00–0.71)      | 0.23 (0.1–0.5)      | 0.09 (0.03–0.3)          | 0.24 (0.2–0.4)         | 0.53 (0.4–0.8)      | 0.22 (0.13–0.36)          |                |
| DOR          | 16.7 (3.3–84.0)        | 12.3 (3.1–48.4)     | 53.8 (19.5–148.6)     | 75.4 (4.2–136.7)      | 20.8 (5.6–77.2)     | 51.1 (13.5–194.1)        | 110.0 (38.1–317.6)     | 11.5 (2.8–46.8)     | 33.0 (20.0–56.0)          |                |
| <b>F3</b>    |                        |                     |                       |                       |                     |                          |                        |                     |                           |                |
| Prevalence   | 39.0%                  | 8.5%                | 31.7%                 | 20.9%                 | 39.3%               | 35.6%                    | 16.2%                  | 11.3%               | 24.1%                     |                |

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

|              | Pittsburgh<br>(n = 59) | Incheon<br>(n = 82) | Yokohama<br>(n = 142) | Rochester<br>(n = 58) | Seville<br>(n = 61) | Los Angeles<br>(n = 101) | San Diego<br>(n = 233) | Uppsala<br>(n = 62)  | Meta-analysis<br>(95% CI) | I <sup>2</sup> |
|--------------|------------------------|---------------------|-----------------------|-----------------------|---------------------|--------------------------|------------------------|----------------------|---------------------------|----------------|
| Cut-off, kPa | 3.40 (3.25–3.8)        | 3.10 (2.3–4.0)      | 4.10 (3.20–4.50)      | 3.95 (3.19–5.81)      | 3.72 (3.62–4.08)    | 3.06 (2.7–4.19)          | 3.60 (3.22–3.86)       | 3.31 (3.24–3.80)     | 3.53 (3.40–3.66)          | 41.15%         |
| AUROC        | 0.95 (0.89–0.99)       | 0.88 (0.74–0.99)    | 0.905 (0.84–0.95)     | 0.92 (0.85–0.99)      | 0.88 (0.78–0.97)    | 0.90 (0.84–0.97)         | 0.95 (0.92–0.98)       | 0.99 (0.94–0.99)     | 0.92 (0.90–0.94)          |                |
| Sensitivity  | 87% (67–95%)           | 71% (29–96%)        | 91% (82–98%)          | 82% (48–98%)          | 75% (53–90%)        | 92% (78–98%)             | 87% (72–96%)           | 100% (47–100%)       | 87% (80–92%)              | 24.42%         |
| Specificity  | 94% (81–99%)           | 96% (89–99%)        | 78% (70–86%)          | 89% (77–96%)          | 86% (71–95%)        | 77% (65–86%)             | 94% (90–97%)           | 94% (85–99%)         | 88% (81–94%)              | 82.98%         |
| PLR          | 31.3 (4.5–217.6)       | 17.86 (5.4–59.5)    | 3.74 (2.43–5.76)      | 3.54 (2.2–5.6)        | 9.25 (3.1–28.0)     | 3.97 (2.5–6.3)           | 14.26 (8.1–25.0)       | 23.6 (5.8–95.0)      | 7.8 (4.7–13.0)            |                |
| NLR          | 0.13 (0.05–0.40)       | 0.30 (0.09–1.0)     | 0.12 (0.04–0.33)      | 0.05 (0.004–0.82)     | 0.27 (0.1–0.5)      | 0.11 (0.04–0.33)         | 0.14 (0.06–0.3)        | 0.15 (0.02–0.9)      | 0.14 (0.09–0.22)          |                |
| DOR          | 233.3 (22.7–2,395.6)   | 60.0 (8.1–445.9)    | 37.1 (11.9–115.4)     | 62.0 (3.4–1123.5)     | 34.0 (7.6–152.2)    | 36.7 (9.8–136.6)         | 10.9 (4.1–28.7)        | 315.0 (13.8–7,214.5) | 55.0 (31.0–98.0)          |                |
| <b>F4</b>    |                        |                     |                       |                       |                     |                          |                        |                      |                           |                |
| Prevalence   | 15.3%                  | 1.2%                | 7.7%                  | 12.1%                 | 11.5%               | 17.8%                    | 6.4%                   | 4.8%                 | 8.7%                      |                |
| Cut-off, kPa | 3.46 (3.13–4.45)       | ND                  | 6.40 (5.40–7.20)      | 4.73 (3.33–5.81)      | 3.82 (3.72–5.00)    | 4.39 (3.98–5.51)         | 4.14 (3.22–4.58)       | ND                   | 4.45 (3.67–5.22)          | 85.24%         |
| AUROC        | 0.95 (0.85–0.99)       | ND                  | 0.95 (0.92–0.98)      | 0.94 (0.86–1.00)      | 0.90 (0.81–0.99)    | 0.87 (0.80–0.95)         | 0.95 (0.92–0.98)       | ND                   | 0.94 (0.92–0.96)          |                |
| Sensitivity  | 100% (66–100%)         | ND                  | 82% (48–98%)          | 86% (42–100%)         | 100% (59–100%)      | 72% (47–90%)             | 87% (60–98%)           | ND                   | 88% (71–96%)              | 5.82%          |
| Specificity  | 76% (62–87%)           | ND                  | 95% (89–97%)          | 90% (79–97%)          | 78% (64–88%)        | 89% (80–95%)             | 93% (89–96%)           | ND                   | 89% (82–93%)              | 77.68%         |
| PLR          | 4.05 (2.39–6.84)       | ND                  | 16.50 (7.1–32.6)      | 8.74 (3.6–21.2)       | 4.08 (2.3–7.2)      | 6.66 (3.4–13.1)          | 12.83 (7.9–21.0)       | ND                   | 7.8 (5.1–11.9)            |                |
| NLR          | 0.07 (0.004–0.98)      | ND                  | 0.10 (0.03–0.59)      | 0.16 (0.03–1.00)      | 0.16 (0.03–1.0)     | 0.31 (0.1–0.7)           | 0.07 (0.01–0.5)        | ND                   | 0.13 (0.05–0.36)          |                |
| DOR          | 65.3 (3.5–1,208.4)     | ND                  | 165.1 (17.6–1,344.2)  | 55.2 (5.5–555.8)      | 51.0 (2.7–956.4)    | 21.4 (6.2–74.0)          | 189.5 (23.3–118.2)     | ND                   | 59.0 (21.0–169.0)         |                |

The bootstrap method and maximum Youden index were applied to estimate the optimal cut-off and 95% CI. In each cohort, ROC analysis was used to estimate AUROC, sensitivity, specificity, PLR, NLR, and DOR. The bivariate random-effects model was used for the computation. Heterogeneity was assessed using the Higgins inconsistency index test.

AUROC, area under the ROC curve; DOR, diagnostic odds ratio; ND, not determined (the calculation was not possible owing to the low prevalence); NLR, negative likelihood ratio; PLR, positive likelihood ratio; ROC, receiver operating curve.

**Table 4.**

LMMs explore the variables affecting LSM by MRE.

| Covariates                    | Univariable |                  |         | Multivariable |                  |         |
|-------------------------------|-------------|------------------|---------|---------------|------------------|---------|
|                               | Estimate    | 95% CI           | p value | Estimate      | 95% CI           | p value |
| Age (years)                   | 0.030       | 0.023 to 0.038   | 0.000   | 0.006         | 0.000 to 0.011   | <0.05   |
| ALT (U/L)                     | 0.002       | 0.000 to 0.004   | 0.026   | 0.000         | -0.002 to 0.002  | 0.920   |
| AST (U/L)                     | 0.006       | 0.003 to 0.009   | 0.000   | 0.002         | -0.001 to 0.005  | 0.255   |
| AST/ALT                       | 0.128       | -0.124 to 0.381  | 0.319   |               |                  |         |
| GGT (U/L)                     | 0.004       | 0.002 to 0.005   | 0.000   | 0.001         | 0.000 to 0.002   | <0.01   |
| Platelet (10 <sup>9</sup> /L) | -0.007      | -0.008 to -0.005 | 0.000   | -0.002        | -0.003 to -0.001 | <0.01   |
| Albumin (g/dl)                | -0.777      | -1.028 to -0.527 | 0.000   | -0.043        | -0.223 to 0.137  | 0.641   |
| Glucose (mg/dl)               | 0.004       | 0.002 to 0.006   | 0.001   | 0.001         | -0.001 to 0.002  | 0.303   |
| HDL (mg/dl)                   | 0.000       | -0.006 to 0.006  | 0.906   |               |                  |         |
| LDL (mg/dl)                   | -0.005      | -0.008 to -0.001 | 0.008   | -0.001        | -0.006 to 0.003  | 0.526   |
| Cholesterol (mg/dl)           | -0.003      | -0.006 to 0.000  | 0.049   | 0.002         | -0.001 to 0.006  | 0.227   |
| MRI-PDFF                      | -0.024      | -0.037 to -0.012 | 0.000   | -0.008        | -0.018 to 0.001  | 0.077   |
| Steatosis stage               |             |                  |         |               |                  |         |
| S1-S3 vs. S0                  | 0.044       | -0.627 to 0.715  | 0.898   |               |                  |         |
| S2-S3 vs. S0-S1               | -0.145      | -0.374 to 0.084  | 0.214   |               |                  |         |
| S3 vs. S0-S2                  | 0.020       | -0.301 to 0.341  | 0.903   |               |                  |         |
| NASH (no/MMA/SA)              |             |                  |         |               |                  |         |
| NASH-MMA                      | -0.092      | -0.362 to 0.178  | 0.504   | 0.083         | -0.114 to 0.280  | 0.409   |
| NASH-SA                       | 0.717       | 0.484 to 0.950   | 0.000   | 0.256         | 0.079 to 0.433   | <0.01   |
| Fibrosis stage                |             |                  |         |               |                  |         |
| F1-F4 vs. F0                  | 0.309       | 0.117 to 0.502   | 0.002   | 0.237         | 0.041 to 0.433   | <0.05   |
| F2-F4 vs. F0-F1               | 0.825       | 0.611 to 1.309   | 0.000   | 0.695         | 0.484 to 0.906   | <0.01   |
| F3-F4 vs. F0-F2               | 1.043       | 0.792 to 1.293   | 0.000   | 0.879         | 0.627 to 1.132   | <0.01   |
| F4 vs. F0-F3                  | 1.438       | 1.139 to 1.737   | 0.000   | 1.402         | 1.100 to 1.704   | <0.01   |

LMM, univariable levels of significance,  $p < 0.10$ ; multivariable levels of significance,  $p < 0.05$ .

ALT, alanine aminotransferase; AST, aspartate aminotransferase; GGT, gamma-glutamyltransferase; LMM, linear mixed-effects model; LSM, liver stiffness measurement; MMA, mild-moderate activity; MRE, magnetic resonance elastography; MRI-PDFF, magnetic resonance imaging proton density fat fraction; SA, severe activity.

**Table 5.**

GLMM exploring variables associated with discrepancies (overestimation and underestimation): including clinical, biochemical, and histological variables.

| Covariates                    | Concordance vs. overestimation |              |               |       | Concordance vs. underestimation |         |               |             |         |   |
|-------------------------------|--------------------------------|--------------|---------------|-------|---------------------------------|---------|---------------|-------------|---------|---|
|                               | Univariable                    |              | Multivariable |       | Univariable                     |         | Multivariable |             |         |   |
|                               | OR                             | 95% CI       | p value       | OR    | 95% CI                          | p value | OR            | 95% CI      | p value |   |
| BMI                           |                                |              |               |       |                                 |         |               |             |         |   |
| 30–35 kg/m <sup>2</sup>       | 1.303                          | 0.638–2.659  | 0.467         | 1.134 | 0.533–2.413                     | 0.744   | 0.928         | 0.441–1.953 | 0.844   | — |
| >35 kg/m <sup>2</sup>         | 2.104                          | 0.950–4.662  | 0.067         | 1.775 | 0.764–4.124                     | 0.182   | 0.761         | 0.324–1.790 | 0.531   | — |
| Age                           |                                |              |               |       |                                 |         |               |             |         |   |
| 45–60 years                   | 1.350                          | 0.603–3.025  | 0.466         |       |                                 |         | 0.942         | 0.435–2.041 | 0.879   | — |
| >60 years                     | 1.583                          | 0.710–3.531  | 0.261         |       |                                 |         | 0.719         | 0.304–1.698 | 0.451   | — |
| T2DM (yes/no)                 | 1.931                          | 1.072–3.478  | 0.028         | 1.475 | 0.793–2.743                     | 0.219   | 0.787         | 0.406–1.527 | 0.479   | — |
| ALT                           |                                |              |               |       |                                 |         |               |             |         |   |
| 50–100 U/L                    | 1.239                          | 0.651–2.359  | 0.514         |       |                                 |         | 0.902         | 0.435–1.872 | 0.782   | — |
| >100 U/L                      | 1.984                          | 0.821–4.799  | 0.128         |       |                                 |         | 1.369         | 0.549–3.413 | 0.501   | — |
| AST                           |                                |              |               |       |                                 |         |               |             |         |   |
| 50–100 U/L                    | 2.196                          | 1.177–4.098  | 0.013         | 1.498 | 0.738–3.038                     | 0.263   | 0.747         | 0.331–1.683 | 0.481   | — |
| >100 U/L                      | 2.901                          | 1.145–7.348  | 0.025         | 1.607 | 0.568–4.543                     | 0.371   | 1.425         | 0.466–4.353 | 0.535   | — |
| AST/ALT                       |                                |              |               |       |                                 |         |               |             |         |   |
| 0.6–1.1                       | 2.447                          | 0.916–6.538  | 0.074         | 2.349 | 0.852–6.472                     | 0.099   | 1.230         | 0.550–2.752 | 0.614   | — |
| >1.1                          | 2.541                          | 0.863–7.483  | 0.091         | 2.074 | 0.668–6.435                     | 0.207   | 1.316         | 0.486–3.563 | 0.588   | — |
| GGT                           |                                |              |               |       |                                 |         |               |             |         |   |
| 60–120 U/L                    | 0.996                          | 0.424–2.340  | 0.992         | 0.705 | 0.285–1.749                     | 0.747   | 1.058         | 0.489–2.291 | 0.887   | — |
| >120 U/L                      | 5.075                          | 2.521–10.214 | 0.000         | 3.388 | 1.577–7.278                     | <0.01   | 1.644         | 0.711–3.799 | 0.245   | — |
| Platelet (10 <sup>9</sup> /L) | 0.998                          | 0.994–1.002  | 0.452         |       |                                 |         | 1.000         | 0.996–1.005 | 0.827   | — |
| MRI-PDFF                      | 1.025                          | 0.992–1.059  | 0.134         |       |                                 |         | 1.006         | 0.969–1.045 | 0.759   | — |
| Steatosis stage               |                                |              |               |       |                                 |         |               |             |         |   |
| S2-S3 vs. S0-S1               | 1.049                          | 0.561–1.962  | 0.880         |       |                                 |         | 1.123         | 0.555–2.271 | 0.747   | — |
| S3 vs. S0-S2                  | 1.336                          | 0.557–3.203  | 0.516         |       |                                 |         | 1.067         | 0.428–2.661 | 0.890   | — |
| NASH (no/MMA/SA)              |                                |              |               |       |                                 |         |               |             |         |   |

| Covariates | Concordance vs. overestimation |             |               |       | Concordance vs. underestimation |         |               |             |         |   |   |
|------------|--------------------------------|-------------|---------------|-------|---------------------------------|---------|---------------|-------------|---------|---|---|
|            | Univariable                    |             | Multivariable |       | Univariable                     |         | Multivariable |             |         |   |   |
|            | OR                             | 95% CI      | p value       | OR    | 95% CI                          | p value | OR            | 95% CI      | p value |   |   |
| NASH-MMA   | 1.563                          | 0.651–3.749 | 0.317         | 1.399 | 0.551–3.549                     | 0.480   | 1.798         | 0.768–4.209 | 0.177   | — | — |
| NASH-SA    | 3.210                          | 1.614–6.382 | 0.001         | 2.448 | 1.180–5.079                     | <0.05   | 1.774         | 0.811–3.879 | 0.151   | — | — |

GLMM, univariable levels of significance,  $p < 0.10$ ; multivariable levels of significance,  $p < 0.05$ .

ALT, alanine aminotransferase; AST, aspartate aminotransferase; GGT, gamma-glutamyl transferase; GLMM, generalised linear mixed model; MMA, mild-moderate activity; MRI-PDFF, magnetic resonance imaging proton density fat fraction; OR, odds ratio; SA, severe activity; T2DM, type 2 diabetes mellitus.

**Table 6.**

GLMM exploring variables associated with discrepancies (overestimation and underestimation): multivariate analysis to be implemented in clinical practice.

| Covariates                    | Concordance vs. overestimation |              |         |               |             |         | Concordance vs. underestimation |             |         |               |        |         |
|-------------------------------|--------------------------------|--------------|---------|---------------|-------------|---------|---------------------------------|-------------|---------|---------------|--------|---------|
|                               | Univariable                    |              |         | Multivariable |             |         | Univariable                     |             |         | Multivariable |        |         |
|                               | OR                             | 95% CI       | p value | OR            | 95% CI      | p value | OR                              | 95% CI      | p value | OR            | 95% CI | p value |
| <b>BMI</b>                    |                                |              |         |               |             |         |                                 |             |         |               |        |         |
| 30–35 kg/m <sup>2</sup>       | 1.303                          | 0.638–2.659  | 0.467   | 1.146         | 0.544–2.416 | 0.718   | 0.928                           | 0.441–1.953 | 0.844   | —             | —      | —       |
| >35 kg/m <sup>2</sup>         | 2.104                          | 0.950–4.662  | 0.067   | 1.777         | 0.771–4.094 | 0.177   | 0.761                           | 0.324–1.790 | 0.531   | —             | —      | —       |
| <b>Age</b>                    |                                |              |         |               |             |         |                                 |             |         |               |        |         |
| 45–60 years                   | 1.350                          | 0.603–3.025  | 0.466   | —             | —           | —       | 0.942                           | 0.435–2.041 | 0.879   | —             | —      | —       |
| >60 years                     | 1.583                          | 0.710–3.531  | 0.261   | —             | —           | —       | 0.719                           | 0.304–1.698 | 0.451   | —             | —      | —       |
| T2DM (yes/no)                 | 1.931                          | 1.072–3.478  | 0.028   | 1.558         | 0.842–2.883 | 0.158   | 0.787                           | 0.406–1.527 | 0.479   | —             | —      | —       |
| <b>ALT</b>                    |                                |              |         |               |             |         |                                 |             |         |               |        |         |
| 50–100 U/L                    | 1.239                          | 0.651–2.359  | 0.514   | —             | —           | —       | 0.902                           | 0.435–1.872 | 0.782   | —             | —      | —       |
| >100 U/L                      | 1.984                          | 0.821–4.799  | 0.128   | —             | —           | —       | 1.369                           | 0.549–3.413 | 0.501   | —             | —      | —       |
| <b>AST</b>                    |                                |              |         |               |             |         |                                 |             |         |               |        |         |
| 50–100 U/L                    | 2.196                          | 1.177–4.098  | 0.013   | 1.667         | 0.832–3.339 | 0.149   | 0.747                           | 0.331–1.683 | 0.481   | —             | —      | —       |
| >100 U/L                      | 2.901                          | 1.145–7.348  | 0.025   | 2.067         | 0.749–5.704 | 0.161   | 1.425                           | 0.466–4.353 | 0.535   | —             | —      | —       |
| <b>AST/ALT</b>                |                                |              |         |               |             |         |                                 |             |         |               |        |         |
| 0.6–1.1                       | 2.447                          | 0.916–6.538  | 0.074   | 2.391         | 0.873–6.554 | 0.090   | 1.230                           | 0.550–2.752 | 0.614   | —             | —      | —       |
| >1.1                          | 2.541                          | 0.863–7.483  | 0.091   | 2.047         | 0.658–6.372 | 0.216   | 1.316                           | 0.486–3.563 | 0.588   | —             | —      | —       |
| <b>GGT</b>                    |                                |              |         |               |             |         |                                 |             |         |               |        |         |
| 60–120 U/L                    | 0.996                          | 0.424–2.340  | 0.992   | 0.747         | 0.306–1.824 | 0.521   | 1.058                           | 0.489–2.291 | 0.887   | —             | —      | —       |
| >120 U/L                      | 5.075                          | 2.521–10.214 | 0.000   | 3.700         | 1.749–7.825 | <0.01   | 1.644                           | 0.711–3.799 | 0.245   | —             | —      | —       |
| Platelet (10 <sup>9</sup> /L) | 0.998                          | 0.994–1.002  | 0.452   | —             | —           | —       | 1.000                           | 0.996–1.005 | 0.827   | —             | —      | —       |
| MRI-PDFF                      | 1.025                          | 0.992–1.059  | 0.134   | —             | —           | —       | 1.006                           | 0.969–1.045 | 0.759   | —             | —      | —       |

GLMM, univariable levels of significance,  $p < 0.10$ ; multivariable levels of significance,  $p < 0.05$ .

ALT, alanine aminotransferase; AST, aspartate aminotransferase; GGT, gamma-glutamyl transferase; GLMM, generalised linear mixed model; MRI-PDFF, magnetic resonance imaging proton density fat fraction; OR, odds ratio; T2DM, type 2 diabetes mellitus.