SYSTEMATIC REVIEW AND META-ANALYSIS: NON-INVASIVE DETECTION OF NONALCOHOLIC FATTY LIVER DISEASE RELATED FIBROSIS IN THE OBESE

Geraldine J. Ooi^{1,2}
Sara Mgaith^{3,4}
Guy D. Eslick⁵
Paul R. Burton^{1,2}
William W. Kemp³
Stuart K. Roberts³
Wendy A. Brown^{1,2}

Affiliations:

1. Centre for Obesity Research and Education, Monash University, Melbourne, Australia
2. Department of Surgery, The Alfred, Melbourne, Australia
3. Department of Gastroenterology, The Alfred, Melbourne, Australia
4. Department of Gastroenterology, Royal Melbourne Hospital, Melbourne, Australia
5. The Whiteley-Martin Research Centre, Discipline of Surgery, The University of Sydney, Nepean Hospital, Penrith, Australia

Keywords: nonalcoholic fatty liver disease, obesity, morbid obesity, liver fibrosis, elastography

Short title: Non-invasive tests for NAFLD

Corresponding author:

Geraldine Ooi
Centre for Obesity Research and Education, Monash
University
99 Commercial Road, Melbourne, 3004
Tel: +613 9903 0725
Fax: +613 9903 0717
Email: geraldine.ooi@monash.edu

Abbreviations: NAFLD – non-alcoholic fatty liver disease; BMI – body mass index; NASH – non-alcoholic steatohepatitis; NFS – NAFLD fibrosis score; ELF – enhanced liver fibrosis; TE – transient elastography; SWE – shearwave elastography; MRE – magnetic resonance elastography; ARI – acoustic radiation force impulse; AUROC – area under the receiver operator characteristic; QUADAS-2 – Quality Assessment of Diagnostic Accuracy Studies 2; PPV – positive predictive value; NPV – negative predictive value; DOR – diagnostic odds ratio; CI – confidence interval; SROC – summary receiver operator characteristic; AUC – area under the curve; ALT – alanine aminotransferase; SE – standard error.

Word count: 6387 words (including references) with four figures, three tables and one supplementary file.

Disclosures: GJO, PRB and WAB report being affiliated with the Centre for Obesity Research and Education. The Centre has received funding for research purposes from Allergan and Apollo Endosurgery, the manufacturers of the LapBand™. The grant is not tied to any specific research project, and neither Allergan nor Apollo Endosurgery have control of the protocol, analysis and reporting of any studies. The Centre also receives a grant from Applied Medical towards educational programs.

WAB reports financial support for a bariatric surgery registry from the Commonwealth of Australia, Apollo Endosurgery, Covidien, Johnson and Johnson, Gore and Applied Medical; and a speaker's honorarium from Merck Sharpe and Dohme. These are outside of the submitted work.

GJO reports scholarships from the National Health and Medical Research Council and the Royal Australasian College of Surgeons.

SM, GDE, SKR and WWK have no other disclosures or conflicts of interest.

ABSTRACT

Background: Non-alcoholic fatty liver disease (NAFLD) is a significant disease burden in obesity. Liver fibrosis is an important prognostic factor in NAFLD, and detection is vital. The pathophysiological changes of obesity can alter the accuracy of non-invasive NAFLD tests. We aimed to review current evidence for common non-invasive tests for NAFLD-related fibrosis in obesity.

Methods: We systematically searched for studies assessing the diagnostic accuracy of 11 biomarker panels and elastography techniques for NAFLD-related fibrosis in obesity. Meta-analyses were performed where possible.

Results: Thirty-eight studies were identified assessing the selected tests in obese populations. Simple biomarker panels (e.g. NAFLD fibrosis score (NFS)) were the most validated. Evidence showed better accuracy of complex biomarker panels (NFS: SROC 0.795-0.813 vs. ELF: SROC 0.962), however these were poorly validated in obesity. Elastography techniques were better studied, and had high diagnostic accuracy (transient elastography: SROC 0.859; magnetic resonance elastography: SROC 0.965), but were limited by BMI-dependent failure. Limited evidence was found to validate the accuracy of any test in exclusively obese populations.

Conclusion: In obese subjects, complex biomarker panels and elastography have reasonable to good accuracy for NAFLD-related fibrosis, however these methods have not been well validated. Further study in this high-risk population is needed.

INTRODUCTION

Obesity, especially coupled with insulin resistance, is a significant risk factor for the development and progression of nonalcoholic fatty liver disease (NAFLD). Therefore, it is not surprising that NAFLD is endemic in obese populations, affecting up to 90% of patients with a body mass index (BMI) greater than 30kg/m^2 (1). The more severe form, non-alcoholic steatohepatitis (NASH), is present in 36-67% of patients with obesity, and advanced fibrosis in up to 30% (2, 3).

Liver fibrosis, in particular, is an important prognostic factor for patients with NAFLD. Fibrosis stage, regardless of steatosis or inflammation status, has been shown to be the only independent histological variable associated with overall and liver-related mortality (4). Even early fibrosis (F1) has been significantly associated with death or liver transplantation (hazard ratio 1.88) (4).

Liver biopsy remains the gold standard for diagnosing non-alcoholic fatty liver disease, and staging fibrosis (5). However, it is an invasive procedure with inherent risks including up to 0.1% incidence of significant complications, and 0.01% risk of mortality. Moreover, other major limitations of liver biopsy include inter-observer variability, sampling error, and cost (5). These factors make it a poor screening tool for the sizeable at-risk population.

Non-invasive methods for assessment of liver fibrosis have been used with increasing frequency in clinical practice. The two categories of commonly utilised diagnostic techniques are those based on serum markers and those based on elastography (**Supplementary Table 1** and **Supplementary Section 1**) (6). Most of these modalities were developed and validated

primarily in studies of viral hepatitis patients, however disease specific biomarkers have also been trialed in NAFLD with varying success.

Whilst many of the common diagnostic techniques have been widely tested in general NAFLD populations, fewer studies have focused on the high risk, yet unselected, obese population. This is important, as morbid obesity represents a very different metabolic, biochemical and physical state to general populations (7). In addition, BMI is a particularly important consideration when the feasibility of elastography techniques is assessed, as the accuracy and success rates are lower in those with obesity (8). Furthermore, studies often include populations selected due to abnormal serum aminotransferase levels, which excludes the known significant proportion of patients with liver disease who have aminotransferase levels within the normal range (9).

The objective of this systematic review and meta-analysis was to gather and evaluate the current evidence available on the accuracy of non-invasive techniques for the diagnosis of NAFLD-related fibrosis, focusing on the obese. A detailed description of all chosen tests can be found in the **Supplementary Materials (Supplementary Section 1)**. We aimed to assess tests that are more commonly used in clinical practice, particularly the NAFLD fibrosis score (NFS), Enhanced Liver Fibrosis score (ELF), Transient Elastography (TE) and Magnetic Resonance Elastography (MRE) (6).

METHODS

For this systematic review, we focused on the most commonly available and widely used non-invasive tests (**Supplementary Table 1** and **Supplementary Section 1**) (6) including:

1. Biomarker panels: NAFLD fibrosis score (NFS), enhanced liver fibrosis (ELF) score, BARD score, FIB4 score, FibroMeter, Fibrotest, Hepascore; and 2. Elastography techniques: Transient elastography (TE), magnetic resonance elastography (MRE), shearwave elastography (SWE), acoustic radiation force impulse (ARFI).

Search methodology

Two independent reviewers (GO, SM) performed a comprehensive literature search to identify studies on non-invasive diagnosis of NAFLD-related fibrosis in obese patients, according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement for the conduct of meta-analyses of observational studies. We searched MEDLINE (1946 to November 2016), EMBASE (1974 to November 2016), Science Citation Index (until November 2016), and the Cochrane Library (until November 2016). Our primary search strategy comprised variations of text words as well as MeSH terms for each test of interest, NAFLD (*nonalcoholic fatty liver disease, nonalcoholic steatohepatitis*) AND liver biopsy (*biopsy, histology, histopathology*). The search and study selection was limited to English language. An additional MEDLINE search was performed for existing systematic reviews. Manual searching of reference lists from reviews, as well as references from selected primary studies was performed to identify additional studies.

Study selection and eligibility criteria

For the systematic review studies had to meet all of the following inclusion criteria: a) liver biopsy as the reference; b) assessment of the index test against reference standard; c) ≥ 18

years old; 4) ≥ 20 participants with NAFLD; d) published full-length manuscripts; and e) focus on obesity, with either i) recruitment of patients with $\text{BMI} \geq 30 \text{ kg/m}^2$ (“obese-only” study), or ii) studies with an average $\text{BMI} \geq 30 \text{ kg/m}^2$ (“average-obese” study). Exclusion criteria were: a) non-English based publications; and b) animal studies. Both prospective and retrospective studies were included. Studies with multiple aetiologies were accepted if data were separable. Additionally, studies including patients with and without liver biopsy, but where data for patients with liver biopsies were separable, were included.

Data extraction and meta-analysis

Both reviewers performed data extraction independently. Any disputes were resolved through consensus. Information extracted included study data (year, country, design, study size), demographic data, recruitment setting including selection from exclusively obese (“obese only”) populations, prevalence of fibrosis, and the index tests and their details. Measures of diagnostic accuracy were extracted for each index test, each level of fibrosis diagnosed, and each threshold used. The studies included in the systematic review were scrutinised, and 2x2 tables and the area under the receiver operator curve (AUROC) for each index test were extracted. Where 2x2 tables could not be formed from published information, efforts were made to contact the relevant corresponding authors for study-level data. Corresponding authors were contacted three times via email and if these efforts failed, the studies were excluded from the meta-analysis.

Quality assessment

Review-specific criteria were developed for the QUADAS-2 tool for quality assessment of diagnostic accuracy studies (**Supplementary Table 5**). Two reviewers independently assessed the quality of included studies.

Data analysis

The weighted mean pooled sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), diagnostic odds ratio (DOR), and their 95% confidence intervals (CIs) were calculated using the DerSimonian and Laird random effects model because of anticipated heterogeneity. Random effects modeling takes into account both within-study and between-study variation. To correct for any continuity errors, 0.5 was added to all cells with a frequency of 0 in order to calculate the pooled estimates using the Yates correction method (10). Summary receiver operating characteristics (SROC) curves were fitted using the Moses-Shapiro-Littenberg method, and the area under the receiver operator characteristic curve (AUROC), Q* index, and their respective standard errors were estimated. The Spearman's correlation coefficient was calculated to assess for threshold effect, and variability between individual studies was evaluated by plotting the diagnostic accuracy estimates on a forest plot. Heterogeneity was quantified using the I^2 index, I^2 values of 25, 50 and 75% corresponded to low, moderate and high degrees of heterogeneity respectively. Publication bias was assessed via Deeks funnel plots and associated regression test of asymmetry (11). Statistical analyses were performed using Meta-Disc (version 1.4, Unit of Clinical Biostatistics, Ramon y Cajal Hospital, Madrid, Spain), GraphPad Prism (version 6.0, GraphPad Software, San Diego, CA), and Microsoft Excel (version 14.2.0, Microsoft, 2011).

RESULTS

Literature search

Using the search strategy as described, we identified 1505 studies. After removal of duplicates, we screened titles and abstracts of 1084 studies. Ninety full text articles were assessed for inclusion by the review criteria specified. Fifty-two studies were excluded due to inseparable data from mixed aetiology studies (n=9), and having a mean study BMI<30kg/m² (n=43). Therefore, thirty-eight studies were included in the systematic review (**Figure 1**).

Study characteristics

The characteristics of the included studies are shown in **Table 1**. Study size varied from 28 to 827 patients. Most studies recruited from Liver disease setting (“*average-obese*” study), with only twelve studies recruiting specifically from settings where only obese patients are treated (“*obese-only*” study). These *obese-only* settings included Bariatric or Obesity clinics (7, 8, 24, 27, 30, 35, 36, 38-40, 42, 46). Demographics and baseline characteristics of *obese-only* cohorts were substantially different from *average-obese* cohorts (**Table 2**). *Obese-only* cohorts were larger, generally younger, with a higher proportion of females, lower rates of advanced fibrosis and lower average ALT levels.

The number of studies that assessed each test varied, from the most validated test with 22 studies (NAFLD fibrosis score) to only one study (SWE) (**Supplementary Table 2a**). There were far fewer *obese-only* studies (**Supplementary Table 2b**). Cut-off histological fibrosis level varied, with assessment of *any fibrosis* (F1-4), *significant fibrosis* (F2-4), *advanced fibrosis* (F3-4), and *cirrhosis* (F4). The most commonly assessed fibrosis grade was *advanced fibrosis* (F3-4).

There were two pairs of studies with a proportion of patient overlap (17, 18, 33, 34), however with some data examining different diagnostic techniques. These were included in the systematic review, but data for similar tests and fibrosis levels were excluded in the meta-analysis.

Publication bias was assessed with Deeks funnel plot and associated regression test for asymmetry (11), for tests assessing seven or more studies (**Supplementary Figure 1**). This showed the potential for publication bias in some tests including the BARD ($p=0.009$), FIB4 ($p=0.016$) and transient elastography ($p=0.046$).

Patient recruitment and selection

The included studies consisted of 22 prospective studies, 14 retrospective studies, one mixed prospective/retrospective recruitment and one with unclear recruitment (**Table 1**).

Twenty-three studies assessed diagnostic accuracy in a population with confirmed histological diagnosis of NAFLD, excluding patients with <5% steatosis. Fifteen studies (7, 8, 16, 21, 23-25, 27, 28, 30, 36, 38, 39, 41, 46) included all recruited patients in the final analysis, regardless of histological NAFLD diagnosis. Nine of these recruited patients from *obese-only* cohorts. Five bariatric surgical studies reported liver biopsies on all bariatric surgical patients (7, 27, 30, 38, 39). In all studies, one or two dedicated pathologists scored all liver biopsies. All studies reported using the NASH Clinical Research Network histological criteria and Kleiner or Brunt classification of fibrosis. Common indications for inclusion in many studies were the use of abnormal aminotransferase levels or ultrasound scans for referral to the recruitment setting. All studies reported the exclusion of other causes of liver disease prior to diagnosis of NAFLD. Additional study exclusions were prolonged time interval between biopsy and diagnostic test, or poor quality tests.

Methodological quality and risk of bias in included studies

Quality assessment via the QUADAS-2 criteria is seen in **Figure 2**. Studies not recruited from an *obese-only* population were considered to be of high applicability concern due to inclusion of non-obese patients within the final analysis. Additionally, a significant number of studies used non-conventional threshold values that were not pre-specified nor validated (**Supplementary Figure 2-5**), creating substantial variation in test application between studies. Blinding was specified in 22 studies (7, 8, 12, 15-19, 21, 22, 24, 25, 27, 28, 30, 32, 35, 38, 40, 41, 44, 46).

Diagnostic test accuracy for all included studies

Using available 2x2 tables, forest plots for diagnostic accuracy were created (**Supplementary Figure 2-5**), with pooled meta-analysis results are reported in **Table 3** (sensitivity, specificity and diagnostic odds ratio (DOR)) and **Supplementary Table 3** (positive (LR+) and negative likelihood ratios (LR-)). Summary receiver operator characteristic curves (SROC) shown in **Figure 3, Figure 4 and Supplementary Figure 6**.

The most commonly used and widely available simple biomarker (NAFLD fibrosis score), complex biomarker (ELF) and elastography techniques (TE, MRE) are discussed below.

Similarly, the greatest evidence was available for distinguishing $\geq$ F3 fibrosis (advanced fibrosis), and are discussed in greater detail below. Full details for all other tests and for all other levels of fibrosis are available in the **Supplementary Materials**.

Biomarker panels

NAFLD fibrosis score (NFS)

Twenty-two studies determined the diagnostic accuracy of the NAFLD fibrosis score in obese populations. For $\geq F3$, the AUROC ranged from 0.615-0.850, with sensitivity 23-100% for a low threshold (NFS<-1.455, -1.31 or unstated), and specificity 31-100% for a high threshold (NFS>0.676 or unstated). For studies reporting dual thresholds, the indeterminate fraction varied from 13% (19) to 52% (46), with an average of 34.5%. The pooled sensitivity and specificity for a high threshold were 60.2% (56.7-63.7%) and 92.4% (91.1-93.5%) (**Table 2**). For a low threshold, the sensitivity and specificity were 78.6% (75.5-81.4%) and 62.2% (60.3-64.0%). The heterogeneity for these pooled values was considerable (I^2 : 45.5-94.5%). The pooled area under the SROC curve was 0.813 (standard error 0.053) and 0.795 (0.020) for high and low threshold respectively (**Figure 3**).

The results for $\geq F1$, $\geq F2$ and $F4$, corresponding to mild fibrosis, significant fibrosis and cirrhosis respectively, are seen in **Table 3**.

Enhanced Liver Fibrosis (ELF)

Three studies assessed the diagnostic accuracy of the ELF in the obese. For $\geq F3$, two studies showed an AUROC 0.900-0.973, sensitivity 80-93% and specificity 90-97%. The thresholds used differed significantly (-3.37 to 0.358). Pooled sensitivity and specificity were 84.8% (75.0-91.9%) and 91.6% (87.1-94.8%) (**Table 3**). The SROC was 0.962 (0.040), with good consistency in diagnostic accuracy between these two studies (I^2 : 14.8-31.8%) (**Figure 4**). Pooled results for $\geq F1$ and $\geq F2$ are seen in **Table 3**.

Elastography modalities

Transient elastography (TE)

Nine studies examined transient elastography in the obese with NAFLD. For $\geq F3$, AUROC was 0.831-0.938, with sensitivity 57-100% and specificity 61-90%. The threshold values used varied significantly from 7.6-12.5kPa. Pooled sensitivity, specificity and DOR were 82.7% (78.7-86.2%), 72.1% (68.9-75.2%) and 16.6 (10.1-27.3). The SROC was 0.859 (0.017) with moderate to high heterogeneity (I^2 : 32.5-89.6%) (**Figure 4**). There were three studies assessing detection of F4 showing good accuracy, with AUROC 0.870-0.950, sensitivity 65-100% and specificity 76–91% (threshold 7.9-16.1). Pooled sensitivity and specificity was 81.8% (72.2-89.2%) and 78.3% (74.5-81.8%). The pooled results for $\geq F2$ are seen in **Table 3**.

Failure rate related to obesity was reported in seven studies, with the remainder excluding patients based on failed readings (**Supplementary Table 4**). Myers *et al* reported an overall low failure rate of 1.1% in a mixed aetiology cohort. In this study, failure rates related to BMI, being 4.9% with a BMI $\geq 40\text{kg/m}^2$. Naveau *et al* reported a substantially higher failure rate of 11.1%. Patients with failed readings were significantly larger than those with successful readings (47.5 vs. 42.0, $p < 0.005$). Both studies utilised the XL probe. The difference in failure rate could be accounted for by the difference in average study BMI (30.0 vs. 42.3 kg/m^2).

Magnetic resonance elastography (MRE)

Four studies assessed the diagnostic accuracy of MRE. For $\geq F3$, the AUROC was 0.924–0.957. Variable threshold values (3.62–4.15kPa) were used to calculate a sensitivity 82–95% and specificity 89–93%. The pooled SROC was 0.965 (0.015) (**Figure 4**). Pooled sensitivity and specificity were 87.0% (79.2-92.7%) and 91.5% (88.3-94.1%) giving it excellent diagnostic accuracy for detection of advanced fibrosis, with little heterogeneity (I^2 0.0-

22.7%). The mean BMI in these studies were mainly in the low thirties (31.7-34.8 kg/m²).

Pooled data for $\geq F1$, $\geq F2$ and $\geq F4$ are seen in **Table 3**.

Subgroup analysis of obese-only studies

Only 12 studies done in exclusively obese cohorts (**Supplementary Table 2b**). Pooled analyses for *obese-only* studies were not performed, due to low study numbers. Fibrometer, Hepascore, SWE and MRE were not assessed by any *obese-only* studies.

NAFLD fibrosis score (NFS)

The AUROC for differentiating $\geq F3$ was available for six studies, ranging from 0.615–0.776. Using a high threshold, specificity and sensitivity varied significantly, from 33–98% and 20–91% respectively. A low threshold had sensitivity 87–100% and specificity 24–80%.

ELF

Only one study (n=41) reported the diagnostic accuracy for ELF at a level of $\geq F2$. An optimum cut-off of 9.920 was used, producing a sensitivity of 100% and specificity of 87%.

Transient elastography

Three studies assessed the diagnostic accuracy of TE, however one study had a 2.4% prevalence of $\geq F2$ fibrosis (1 of 41 patients). The two remaining studies had an AUROC of 0.850-0.900, sensitivity 100% and specificity 74% for $\geq F3$ fibrosis (threshold 7.6kPa). For $\geq F2$, the AUROC was 0.810–0.850, with sensitivity 73–81% and specificity 65–78% (threshold 6.4–7.6kPa).

DISCUSSION

The aim of our systematic review was to evaluate the accuracy of commonly used non-invasive diagnostic tests for NAFLD-related fibrosis in the high-risk obese population. With limited evidence available, complex composite scores comprising fibrosis biomarkers (such as the ELF) appeared more accurate than simple composite scores (such as the NFS). Elastography techniques had the best and most consistent accuracy, but were limited by failure rates dependent on BMI. Overall, we found that non-invasive tests were not well validated in the obese, with limited studies and high heterogeneity.

Whilst there have been previous reviews on non-invasive tests for NAFLD-related fibrosis (48, 49), there are few meta-analyses published. Specifically, there is limited review of the evidence in the morbidly obese populations, where NAFLD is endemic (2, 50). A meta-analysis on TE by Kwok *et al* reported reasonable accuracy for NAFLD-related fibrosis, but included studies with significant proportions of non-obese individuals (49). A focus on the use of XL probes and failure rates were not included, but are pertinent considerations in high BMI populations.

There are few studies that cover the gap in knowledge about applicability of these tests in the obese. Our systematic review identified only 38 studies assessing the most common and clinically available diagnostic tests. There was significant heterogeneity of results, and the suggestion of possible publication bias. There was substantially less evidence when assessing an *obese-only* population, with only twelve such studies available. Notably, the diagnostic accuracy of non-invasive tests was not as favorable in *obese-only* cohorts. Importantly, there were significant differences between *obese-only* and *average-obese/liver clinic* studies,

particularly in terms of body mass index and ALT levels. This raises the possibility of two distinct groups with differential hepatic risk profiles. As many serum-based scores are derived from the liver clinic setting, it is possible that currently available studies may not adequately reflect the accuracy of these tests in a general morbidly obese population (7).

Of the serum-based diagnostic tests, we found that complex scores show better accuracy than simple scores. Previous individual studies support these findings in general NAFLD populations (12, 15, 20). In our meta-analysis, both the NFS and BARD had average, and often poor, pooled sensitivities and specificities, with high heterogeneity. Furthermore, 35% of patients had indeterminate NFS scores, which considerably affects the clinical efficacy of this scoring system. The assessment of ELF was more favorable and less variable. Certainly, ELF is now used with increasing confidence in many centres, with 2015 NICE guidelines on NAFLD stating that ELF was “the most cost-effective and the most appropriate test for advanced fibrosis in adults with NAFLD” (51). However, consideration should be given to the level of evidence available for the ELF in obese populations, with only two such studies found. Additionally, both studies used variable optimal thresholds to calculate diagnostic accuracy. Further validation of complex serum scores in the obese is necessary before these findings can be confirmed with confidence.

Elastography techniques showed consistently good diagnostic accuracy. The SROC for TE, ARFI and MRE for all levels of fibrosis ranged from 0.851-0.965. Furthermore, pooled sensitivity and specificity of TE and MRE were good to excellent, with little heterogeneity. These results are consistent with previous data suggesting that elastography has advantages over serum-based predictive scores (14, 15). However, one of the current major difficulties

with these techniques, and a barrier to use in clinical practice, is the heterogeneity of cut-off points quoted in the literature. This variation can be seen in **Supplementary Figures 2-5**, showing anywhere from mild variation in MRE thresholds (3.62-4.15kPa) to significant variation in TE (7.6-12.5kPa). These differences may potentially be influenced by the levels of obesity and variation in skin-to-capsule distance, which are known to affect other elastography-based techniques (52, 53).

The success of elastography techniques was limited by patient size. Naveau *et al* found that the 26 patients with unreliable or unfeasible readings had a significantly higher BMI (47 vs. 42, $p<0.005$) (8). Other elastography studies, including those utilizing ARFI and SWE (16, 17, 27) consistently report similar difficulties with increasing BMI (**Supplementary Table 4**). Myers *et al* reported a failure rate consistent with BMI, with over half of patients with $BMI \geq 40$ having unsuccessful readings with the M probe. This was partially mitigated by the XL probe, decreasing this to a 5% failure rate in this population (35). The XL probe is relatively new and has been designed specifically for larger subjects. There are currently few published studies on the diagnostic accuracy of TE using this probe in the obese (**Supplementary Table 4**). Further investigation into the impact of the XL probe on feasibility and accuracy could significantly increase the applicability of TE in an obese cohort.

The accuracy and structural information obtained by MRI has always been offset by the practicalities of imaging in obesity (54). Diagnostic and research MRI machines have a weight limit of 250kg, but more importantly, a maximum aperture diameter of up to 70cm. In

addition, image quality can be compromised due to increased sound-to-noise ratio and artifact. This clearly restricts its use in morbid obesity.

The evaluation of non-invasive diagnostic tests for NAFLD, particularly in the obese cohort, is becoming increasingly important. With the rising rates of obesity, metabolic disease and deleterious lifestyle choices, NAFLD and steatofibrosis is becoming an increasingly common cause of end-stage liver disease and liver transplantation (55). Viable treatment options are now available for all stages of NAFLD, and monitoring for liver failure and hepatocellular carcinoma is essential for those with more significant fibrosis. Weight loss is central to treatment of NAFLD in the morbidly obese, with persuasive evidence of regression in disease with weight loss, particularly in those with less severe disease progression (56, 57). Effective and durable weight loss, however, is a notoriously challenging endeavour (58). Lifestyle changes alone are often ineffective in the long term (59), and significant sustained weight loss requires a multidisciplinary step-up approach involving allied health, pharmacology and occasionally, surgery (58). The current lack of reliable, safe and practical diagnostic test presents a significant barrier to initiating this pathway of treatment for NAFLD in the setting of obesity.

There are several strengths of this systematic review. Firstly, our focus on the obese population provides a comprehensive summary of the evidence for this high-risk population. Increasing rates of morbid obesity now affect all medical fields, from primary practice to tertiary referral centres. Importantly, this review has identified the paucity of evidence available in this domain. Diagnosis of significant NAFLD in the obese in the absence of liver biopsy remains challenging, and further research is required. Secondly, this review covers the

most commonly used non-invasive methods of detecting fibrosis, allowing for comparison of results. Lastly, we have chosen to focus on studies with liver biopsy as the reference standard. Although there are drawbacks to liver biopsy (5), it is the accepted gold standard that allows a standardised assessment of NAFLD-related fibrosis.

This review has some limitations that warrant discussion. Firstly, there are a variety of thresholds used across the studies for each test (**Supplementary Figure 2-5**). Whilst long established scores (e.g. the NFS and BARD) are more standardised, newer tests are more variable, with studies often report diagnostic accuracy based on optimal thresholds. Unfortunately, considerable variability in threshold values does not allow us to draw practical recommendations from our meta-analysis regarding the best application and interpretation of tests. Secondly, the inclusion of studies based on average BMI $\geq$ 30 has meant that a proportion of data from the meta-analyses includes results from non-obese individuals. These studies include a proportion of non-obese participants, and thus may not completely represent results in an obese cohort. Endeavours to obtain study-level data have been unsuccessful, with poor responses for clarification of study data. Additionally, no studies published analyses of the differences between obese and non-obese subgroups within their studies. There are only twelve studies that have recruited from obese-only populations, with insufficient data for meaningful pooled analyses. Thirdly, although we have assessed publication bias, funnel plots and tests for publication bias have weaknesses in the setting of highly variable results, and may be influenced by type of population and study quality (11, 60). Since variability in test accuracy is expected in diagnostic accuracy studies, the Cochrane Handbook as warned against interpreting statistical evidence of funnel plot asymmetry as necessarily implying publication bias (60). Lastly, we have focused on full-length publications and excluded abstracts, non-English articles, letters, editorials and grey

literature for practical reasons, particularly in relation to having access to sufficient data on diagnostic test accuracy. This may ultimately have had some effect on the results.

In conclusion, this systematic review and meta-analysis focused on the accuracy of established and widely used non-invasive methods in diagnosing NAFLD-related fibrosis in the obese. Complex serum scores, particularly the ELF, had better accuracy compared to simple composite scores, but significantly less validation. Elastography techniques showed the highest accuracy, however, studies report failure rates directly related to BMI. Overall, there is currently limited evidence for non-invasive tests for NAFLD-related fibrosis in the obese, and further studies should be performed to establish their accuracy in this high-risk population.

References

- 1 Bellentani S, Marino M. Epidemiology and natural history of non-alcoholic fatty liver disease (NAFLD). *Ann Hepatol.* 2009; 8:S4-S8.
- 2 Abrams GA, Kunde SS, Lazenby AJ, Clement R. Portal fibrosis and hepatic steatosis in morbidly obese subjects: a spectrum of nonalcoholic fatty liver disease. *Hepatology.* 2004; 40:475-83.
- 3 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.
- 4 Angulo P, Kleiner DE, Dam-Larsen S, Adams L, Bjornsson ES, Charatcharoenwittaya P, *et al.* Liver fibrosis, but no other histologic features, is associated with long-term outcomes of patients with nonalcoholic fatty liver disease. *Gastroenterology.* 2015; 149:389-97.
- 5 Rockey DC, Caldwell SH, Goodman ZD, Nelson RC, Smith AD. Liver biopsy - AASLD Position Paper. *Hepatology.* 2009; 49:1017-44.
- 6 Asrani SK. Incorporation of non-invasive measures of liver fibrosis into clinical practice: diagnosis and prognosis. *Clin Gastroenterol Hepatol.* 2015; 13:2190-204.
- 7 Ooi GJ, Burton PR, Doyle L, Wentworth JM, Bhathal PS, Sikaris K, *et al.* Modified thresholds for fibrosis risk scores in nonalcoholic fatty liver disease are necessary in the obese. *Obes Surg.* 2016; 21(1):115-125.
- 8 Naveau S, Lamouri K, Pourcher G, Njiké-Nakseu M, Ferretti S, Courie R, *et al.* The diagnostic accuracy of transient elastography for the diagnosis of liver fibrosis in bariatric surgery candidates with suspected NAFLD. *Obes Surg.* 2014; 24:1693-701.

- 9 Prati D, Taioli E, Zanella A, Torre ED, Butelli S, Vecchio ED, *et al.* Updated definitions of healthy ranges for serum alanine aminotransferase levels. *Ann Intern Med.* 2002; 137:1-10.
- 10 Walter SD, Cook RJ. A comparison of several point estimators of the odds ratio in a single 2 x 2 contingency table. *Biometrics.* 1991; 47:795-811.
- 11 Deeks JJ, Macaskill P, Irwig L. The performance of tests of publication bias and other sample size effects in systematic reviews of diagnostic test accuracy was assessed. *J Clin Epidemiol.* 2005; 58:882-93.
- 12 Adams LA, George J, Bugianesi E, Rossi E, De Boer WB, van der Poorten D, *et al.* Complex non-invasive fibrosis models are more accurate than simple models in non-alcoholic fatty liver disease. *J Gastroenterol Hepatol.* 2011; 26:1536-43.
- 13 Angulo P, Hui JM, Marchesini G, Bugianesi E, George J, Farrell GC, *et al.* The NAFLD fibrosis score: A noninvasive system that identifies liver fibrosis in patients with NAFLD. *Hepatology.* 2007; 45:846-54.
- 14 Aykut UE, Akyuz U, Yesil A, Eren F, Gerin F, Ergelen R, *et al.* A comparison of FibroMeter NAFLD Score, NAFLD fibrosis score, and transient elastography as noninvasive diagnostic tools for hepatic fibrosis in patients with biopsy-proven non-alcoholic fatty liver disease. *Scand J Gastroenterol.* 2014; 49:1343-8.
- 15 Boursier J, Vergniol J, Guillet A, Hiriart JB, Lannes A, Le Bail B, *et al.* Diagnostic accuracy and prognostic significance of blood fibrosis tests and liver stiffness measurement by Fibroscan in nonalcoholic fatty liver disease. *J Hepatol.* 2016; 65:570-78.
- 16 Cassinotto C, Boursier J, de Ledinghen V, Lebigot J, Lapuyade B, Cales P, *et al.* Liver stiffness in nonalcoholic fatty liver disease: A comparison of supersonic shear imaging, FibroScan, and ARFI with liver biopsy. *Hepatology.* 2016; 63:1817-27.

- 17 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; 41:1271-80.
- 18 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; 63:453-61.
- 19 Demir M, Lang S, Schlattjan M, Drebber U, Wedemeyer I, Nierhoff D, *et al.* NIKEL: a new inexpensive and non-invasive scoring system to exclude advanced fibrosis in patients with NAFLD. *PLoS ONE.* 2013; 8:e58360.
- 20 Dincses E, Yilmaz Y. Diagnostic usefulness of FibroMeter VCTE for hepatic fibrosis in patients with nonalcoholic fatty liver disease. *Eur J Gastroenterol Hepatol.* 2015; 27:1149-53.
- 21 Dvorak K, Stritesky J, Petrtyl J, Vitek L, Sroubkova R, Lenicek M, *et al.* Use of non-invasive parameters of non-alcoholic steatohepatitis and liver fibrosis in daily practice - An exploratory case-control study. *PLoS ONE.* 2014; 9:e111551.
- 22 Ergelen R, Yilmaz Y, Asedov R, Celikel C, Akin H, Bugdayci O, *et al.* Comparison of Doppler ultrasound and transient elastography in the diagnosis of significant fibrosis in patients with nonalcoholic steatohepatitis. *Abdominal radiology (New York).* 2016; 41:1505-10.
- 23 Ergelen R, Akyuz U, Aydin Y, Eren F, Yilmaz Y. Measurements of serum procollagen-III peptide and M30 do not improve the diagnostic accuracy of transient elastography for the detection of hepatic fibrosis in patients with nonalcoholic fatty liver disease. *Eur J Gastroenterol Hepatol.* 2015; 27:667-71.

- 24 Francque SMA, Verrijken A, Mertens I, Hubens G, Van Marck E, Pelckmans P, *et al.* Noninvasive Assessment of Nonalcoholic Fatty Liver Disease in Obese or Overweight Patients. *Clin Gastroenterol Hepatol.* 2012; 10:1162-68.
- 25 Guha IN, Parkes J, Roderick P, Chattopadhyay D, Cross R, Harris S, *et al.* Noninvasive markers of fibrosis in nonalcoholic fatty liver disease: Validating the European Liver Fibrosis Panel and exploring simple markers. *Hepatology.* 2008; 47:455-60.
- 26 Harrison SA, Oliver D, Arnold HL, Gogia S, Neuschwander-Tetri BA. Development and validation of a simple NAFLD clinical scoring system for identifying patients without advanced disease. *Gut.* 2008; 57:1441-7.
- 27 Karlas T, Dietrich A, Peter V, Wittekind C, Lichtinghagen R, Garnov N, *et al.* Evaluation of Transient Elastography, Acoustic Radiation Force Impulse Imaging (ARFI), and Enhanced Liver Function (ELF) Score for Detection of Fibrosis in Morbidly Obese Patients. *PLoS ONE.* 2015; 10:e0141649.
- 28 Kim D, Kim WR, Talwalkar JA, Kim HJ, Ehman RL. Advanced fibrosis in nonalcoholic fatty liver disease: noninvasive assessment with MR elastography. *Radiology.* 2013; 268:411-9.
- 29 Kruger FC, Daniels CR, Kidd M, Swart G, Brundyn K, van Rensburg C, *et al.* APRI: a simple bedside marker for advanced fibrosis that can avoid liver biopsy in patients with NAFLD/NASH. *S Afr Med J.* 2011; 101:477-80.
- 30 Lassailly G, Caiazzo R, Hollebecque A, Buob D, Leteurtre E, Arnalsteen L, *et al.* Validation of noninvasive biomarkers (FibroTest, SteatoTest, and NashTest) for prediction of liver injury in patients with morbid obesity. *Eur J Gastroenterol Hepatol.* 2011; 23:499-506.

- 31 Lee TH, Han SH, Yang JD, Kim D, Ahmed M. Prediction of advanced fibrosis in nonalcoholic fatty liver disease: An enhanced model of BARD score. *Gut Liver*. 2013; 7:323-28.
- 32 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; 60:1920-8.
- 33 McPherson S, Stewart SF, Henderson E, Burt AD, Day CP. Simple non-invasive fibrosis scoring systems can reliably exclude advanced fibrosis in patients with non-alcoholic fatty liver disease. *Gut*. 2010; 59:1265-9.
- 34 McPherson S, Anstee QM, Henderson E, Day CP, Burt AD. Are simple noninvasive scoring systems for fibrosis reliable in patients with NAFLD and normal ALT levels? *Eur J Gastroenterol Hepatol*. 2013; 25:652-8.
- 35 Myers RP, Pomier-Layrargues G, Kirsch R, Pollett A, Duarte-Rojo A, Wong D, *et al*. Feasibility and diagnostic performance of the FibroScan XL probe for liver stiffness measurement in overweight and obese patients. *Hepatology*. 2012; 55:199-208.
- 36 Nassif AT, Nagano TA, Okayama S, Nassif LS, Branco Filho A, Sampaio Neto J. Performance of the Bard Scoring System in Bariatric Surgery Patients with Nonalcoholic Fatty Liver Disease. *Obes Surg*. 2017; 27(2):394-98.
- 37 Palmeri ML, Wang MH, Rouze NC, Abdelmalek MF, Guy CD, Moser B, *et al*. Noninvasive evaluation of hepatic fibrosis using acoustic radiation force-based shear stiffness in patients with nonalcoholic fatty liver disease. *J Hepatol*. 2011; 55:666-72.
- 38 Pimentel SK, Strobel R, Goncalves CG, Sakamoto DG, Ivano FH, Coelho JC. Evaluation of the nonalcoholic fat liver disease fibrosis score for patients undergoing bariatric surgery. *Arq Gastroenterol*. 2010; 47:170-3.

- 39 Praveenraj P, Gomes RM, Basuraju S, Kumar S, Senthilnathan P, Parathasarathi R, *et al.* Preliminary Evaluation of Acoustic Radiation Force Impulse Shear Wave Imaging to Detect Hepatic Fibrosis in Morbidly Obese Patients Before Bariatric Surgery. *J Laparoendosc Adv Surg Tech A.* 2016; 26:192-5.
- 40 Qureshi K, Clements RH, Abrams GA. The utility of the "NAFLD fibrosis score" in morbidly obese subjects with NAFLD. *Obes Surg.* 2008; 18:264-70.
- 41 Ratziu V, Massard J, Charlotte F, Messous D, Imbert-Bismut F, Bonyhay L, *et al.* Diagnostic value of biochemical markers (FibroTest-FibroSURE) for the prediction of liver fibrosis in patients with non-alcoholic fatty liver disease. *BMC Gastroenterol.* 2006; 6:6.
- 42 Rodriguez RD, Pomar MDB, Fernandez AC, Francisco TGd, Goni JL. Usefulness of an index score as a predictor of hepatic fibrosis in obese patients undergoing bariatric surgery. *Rev Esp Enferm Dig.* 2009; 11:528-35.
- 43 Ruffillo G, Fassio E, Alvarez E, Landeira G, Longo C, Dominguez N, *et al.* Comparison of NAFLD fibrosis score and BARD score in predicting fibrosis in nonalcoholic fatty liver disease. *J Hepatol.* 2011; 54:160-3.
- 44 Shah AG, Lydecker A, Murray K, Tetri BN, Contos MJ, Sanyal AJ. Comparison of noninvasive markers of fibrosis in patients with nonalcoholic fatty liver disease. *Clin Gastroenterol Hepatol.* 2009; 7:1104-12.
- 45 Siddiqui MS, Patidar KR, Boyett S, Luketic VA, Puri P, Sanyal AJ. Performance of non-invasive models of fibrosis in predicting mild to moderate fibrosis in patients with non-alcoholic fatty liver disease. *Liver Int.* 2016; 36:572-79.
- 46 Simo KA, McKillop IH, McMillan MT, Ahrens WA, Walters AL, Thompson KJ, *et al.* Does a calculated "NAFLD fibrosis score" reliably negate the need for liver biopsy in patients undergoing bariatric surgery? *Obes Surg.* 2014; 24:15-21.

- 47 Subasi CF, Aykut UE, Yilmaz Y. Comparison of noninvasive scores for the detection of advanced fibrosis in patients with nonalcoholic fatty liver disease. *Eur J Gastroenterol Hepatol*. 2015; 27:137-41.
- 48 Festi D, Marzi SL, Di Biase AR, Mandolesi D, Montrone L, Scaioli E, *et al*. Review article: the diagnosis of non-alcoholic fatty liver disease - availability and accuracy of non-invasive methods. *Aliment Pharmacol Ther*. 2013; 37:392-400.
- 49 Kwok R, Tse YK, Wong GL, Ha Y, Lee AU, Ngu MC, *et al*. Systematic review with meta-analysis: non-invasive assessment of non-alcoholic fatty liver disease: the role of transient elastography and plasma cytokeratin-18 fragments. *Aliment Pharmacol Ther*. 2014; 39:254-69.
- 50 Woo Baidal JA, Lavine JE. The intersection of nonalcoholic fatty liver disease and obesity [Review]. *Sci Transl Med*. 2016; 8:1-12.
- 51 National Institute of Health and Care Excellence. Non-alcoholic fatty liver disease: Assessment and management of non-alcoholic fatty liver disease (NAFLD) in adults, children and young people. NICE guideline - Methods, evidence and recommendations. National Institute for Health and Care Excellence: 2015.
- 52 Shen F, Zheng R-D, Shi J-P, Mi Y-Q, Chen G-F, Hu Z, *et al*. Impact of skin capsular distance on the performance of controlled attenuation parameter in patients with chronic liver disease. *Liver Int*. 2015; 35:2392-400.
- 53 Huang Z, Zheng J, Zeng J, Wang X, Wu T, Zheng R. Normal liver stiffness in healthy adults assessed by real-time shear wave elastography and factors that influence this method. *Ultrasound Med Biol*. 2014; 40:2549-55.
- 54 Uppot RN, Shahani DV, Hahn PF, Gervais D, Mueller PR. Impact of obesity on medical imaging and image-guided intervention. *Am J Roentgenol*. 2007; 188:433-40.

- 55 LaBrecque D, Abbas Z, Anania F, Ferenci P, Khan AG, Goh K-L, *et al.* World Gastroenterology Organisation global guidelines: Nonalcoholic fatty liver disease and nonalcoholic steatohepatitis. *J Clin Gastroenterol.* 2012; 48:467-73.
- 56 Chavez-Tapia NC, Tellez-Avila FI, Barrientos-Gutierrez T, Mendez-Sanchez N, Lizardi-Cervera J, Uribe M. Bariatric surgery for non-alcoholic steatohepatitis in obese patients. *Cochrane Database Syst Rev.* 2010:CD007340.
- 57 Finelli C, Tarantino G. Have guidelines addressing physical activity been established in nonalcoholic fatty liver disease? *World J Gastroenterol.* 2012; 18:6790-800.
- 58 World Health Organisation. Obesity: Preventing and managing the global epidemic. Report of a WHO consultation. WHO technical Report Series: Geneva, Switzerland 2000.
- 59 Sjostrom L, Peltonen M, Jacobson P, Sjostrom CD, Karason K, Wedel H, *et al.* Bariatric surgery and long-term cardiovascular events. *J Am Med Assoc.* 2012; 307:56-65.
- 60 Macaskill P, Gatsonis C, Deeks JJ, Harbord RM, Takwoingi Y. Chapter 10: Analysing and presenting results. In: Deeks JJ, Bossuyt PMM, Gatsonis C (eds.). *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy.* The Cochrane Collaboration 2010.

CONFLICTS OF INTEREST STATEMENT

GJO, PRB and WAB report being affiliated with the Centre for Obesity Research and Education. The Centre has received funding for research purposes from Allergan and Apollo Endosurgery, the manufacturers of the LapBand™. The grant is not tied to any specific research project, and neither Allergan nor Apollo Endosurgery have control of the protocol, analysis and reporting of any studies. The Centre also receives a grant from Applied Medical towards educational programs.

WAB reports financial support for a bariatric surgery registry from the Commonwealth of Australia, Apollo Endosurgery, Covidien, Johnson and Johnson, Gore and Applied Medical; and a speaker's honorarium from Merck Sharpe and Dohme. These are outside of the submitted work.

GJO reports scholarships from the National Health and Medical Research Council and the Royal Australasian College of Surgeons.

SM, GDE, SKR and WWK have no other disclosures or conflicts of interest.

FIGURE LEGEND

Figure	Caption
<i>Figure 1</i>	Study selection flowchart.
<i>Figure 2</i>	QUADAS-2 quality assessment of included studies.
<i>Figure 3</i>	Summary receiver operator curves for detection of F3-4 fibrosis with simple composite serum panels. <i>SROC – summary receiver operator characteristic curve</i>
<i>Figure 4</i>	Summary receiver operator curves for detection of F3-4 fibrosis with complex composite serum panels and elastography techniques. <i>SROC – summary receiver operator characteristic curve</i>
<i>Table 1</i>	Characteristics of included studies.
<i>Table 2</i>	Comparison of obesity-only studies and studies recruiting from liver disease settings.
<i>Table 3</i>	Meta-analysis with pooled sensitivity, specificity and diagnostic odds ratio (DOR), and summary receiver operator characteristic (SROC) curve for detection of each level of fibrosis.

OBR_12628_F1.tif

OBR_12628_F3.tif

OBR_12628_F4.tif

Table 1: *Characteristics of included studies.*

Author, year	Setting	Study type	Country	n=	Age	%male	BMI	%F3/4	ALT	Tests
Adams 2011 [12]	Liver clinic	Prospective	Australia, Italy	242	46.8±12.4	60.3%	30.2	21.9%	66.5	BARD, FIB4, Hepascore, Fibrotest NFS
Angulo 2007 [13]	Liver clinic	Prospective	US, Italy, UK, Australia	733	47.7±13.2	53%	32.2	27.2%	87±72	
Aykut 2014 [14]	Liver clinic	Prospective	Turkey	88	46±9	56.8%	30.3	30.7%	84±56	NFS, Fibrotest, TE
Boursier 2016 [15]	Liver clinic	Prospective	France	452	55.9±12	60%	31.1	38.0%	68±39	NFS, BARD, FIB4, Hepascore, Fibrotest, Fibrometer, TE
Cassinotto 2016 [16]	Liver clinic	Prospective	France	291	56.7±12	59.1%	32.1	43.3%	71.2±50.7	TE, ARFI
Cui 2015 [^] [17]	Liver clinic	Prospective	USA	102	51.3±14	41.2%	31.7	18.6%	58.0±56.1	NFS, BARD, FIB4, MRE
Cui 2016 [18]	Liver clinic	Prospective	USA	125	48.9±15.4	45.6%	31.8	16.8%	56.4±51.8	ARFI, MRE
Demir 2013 [19]	Liver clinic	Prospective	Italy	267	43.8±21.1	47.2%	37	8.2%	56.6±55.2	NFS, BARD, FIB4
Dincses 2015 [20]	Liver clinic	Retrospective	Turkey	52	45±9	57.7%	30.8	19.2%	89±58	NFS, Fibrometer, TE
Dvorak 2014 [21]	Liver clinic	Prospective	Turkey	56	48.9±14.9	70.5%	31.2	30.4%	135±96	NFS, BARD, FIB4, ELF
Ergelen 2015 [22]	Liver clinic	Prospective	Turkey	87	45.8±9	49.4%	30.6	21.8%	77.8±56.1	TE
Ergelen 2016 [23]	Liver clinic	Prospective	Turkey	63	47.1±8.4	61.9%	30.4	50.8% (≥F2)	-	TE
*Francque 2012 [24]	Obesity clinic	Prospective	Belgium	313	43.5±12.7	28.6%	38.6	7.2%	43.3±22.1	NFS, BARD, FIB4
Guha 2008 [25]	Liver clinic	Prospective	UK	196	48.7±12.5	64%	32.4	23%	77.3±57.8	ELF
Harrison 2008 [26]	Liver clinic	Retrospective	USA	827	49 (17-95)	49%	33	30%	69	NFS, BARD
*Karlas 2015 [27]	Bariatric surgical	Prospective	Germany	41	45.7±10.2	32%	46.8	2% (≥F2)	-	ELF, TE, ARFI
Kim 2013 [28]	Liver clinic	Retrospective	USA	142	52.8±12.8	40.3%	34.8	32.4%	70.6±61.0	NFS, FIB4, BARD, MRE
Kruger 2011 [29]	Liver clinic	Uncertain	South Africa	111	52 (50 – 54)	27%	35	17%	-	NFS
*Lassailly 2011 [30]	Bariatric surgical	Prospective	USA	288	41.6 ±12.8	23.6%	48.6	2.4%	34±23	Fibrotest
Lee 2013 [31]	Liver clinic	Retrospective	USA	107	48.9 (40.9–50)	38.3%	35.9	31.8%	63 (29-105)	BARD
Loomba 2014 [32]	Liver clinic	Prospective	USA	117	50.1±13.4	43.6%	32.4	18.8%	66.3±54.4	MRE
McPherson 2010 [^] [33]	Liver clinic	Prospective	UK	145	49.3	61%	35	19%	94±63	NFS, BARD, FIB4
McPherson 2013 [34]	Liver clinic	Retrospective	UK	305	51±12	56%, 63%	35	24%, 17%	28±9, 95±62	NFS, BARD, FIB4
*Myers 2012 [35]	Liver clinic, obese only	Prospective	Canada	75	50 (43-57)	63%	30	29.3%	55 (36–87)	TE
*Nassif 2016 [36]	Bariatric surgical	Retrospective	Brazil	298	40.1	11.1%	43.6	91.5% (≥F1)	-	BARD
*Naveau 2014 [8]	Bariatric surgical	Prospective	France	100	42.5 (SEM 0.5)	19%	42.3	9%	38±29.5	TE
*Ooi 2016 [7]	Bariatric surgical	Prospective	Australia	101	49 (37-54.5)	33.7%	41.9	3.0%	32 (22-42)	NFS, BARD, FIB4
Palmeri 2011 [37]	Liver clinic	Mix	USA	135	-	37.8%	>30	29.6%	-	ARFI
*Pimentel 2010 [38]	Bariatric surgical	Retrospective	Brazil	158	36±10	22.8%	41	14%	55 (36–87)	NFS
*Praveenraj 2016 [39]	Bariatric surgical	Prospective	India	28	48.3	-	49.6	28.6%	21.9	ARFI
*Qureshi 2008 [40]	RYGB patients	Retrospective	USA	331	40.5±9	17%	48.5	13.6%	29.1±15.7	NFS
Ratzu 2006 [41]	Liver, inpatient	Prospective	France	267	51.2	58%	>27	18.8%	71±3, 79±5	Fibrotest

*Rodriguez 2009 [42]	Bariatric surgical	Retrospective	France	88	40.6±11.3	NA	52.7	5.5%	34.3±28.8	NFS
Ruffillo 2011 [43]	Liver clinic	Retrospective	Argentina	138	49 (38-57)	48.6%	30.3	26.8%	69 (50-96.5)	NFS, BARD
Shah 2009 [44]	Liver clinic	Retrospective	USA	541	48±12	40%	34	23.1%	57 (35-83), 68 (47-103)	NFS, BARD, FIB4
Siddiqui 2016 [45]	Liver clinic	Retrospective	USA	145	52.9±11.7	36.7%	35.8	35.2%	80.7±53.3	NFS, BARD, FIB4, Fibrometer
*Simo 2014 [46]	RYGB patients	Retrospective	USA	225	43.2±9.6	14.7%	44.6	6.6%	31.2±24.3	NFS
Subasi 2015 [47]	Liver clinic	Retrospective	Turkey	142	45±9	52.8%	30.9	21.1%	91±61	NFS, BARD, FIB4, Fibrometer

*Obese-only study. ^Some data not included in meta-analyses due to patient overlap with subsequent study. NS

– not stated; NFS – NAFLD fibrosis score, SEM – standard error of the mean; ELF – enhanced liver fibrosis score; TE – transient elastography; MRE – magnetic resonance elastography; ARFI – acoustic radiation force impulse; RYGB – Roux-en-Y gastric bypass

Table 2: Comparison of obesity-only studies and studies recruiting from liver disease settings

	“Average-obese” study (e.g. Liver clinic)	“Obese-only” study (e.g. Obesity clinic)	p-value
n=	26	12	
Age	49.3±3.1 years	43.3±4.0 years	<0.001
Male gender	51.1±10.3%	26.6±14.0%	<0.001
Body mass index	32.4±2.3 kg/m ²	44.0±5.7 kg/m ²	<0.001
Advanced fibrosis (F3-F4)	25.7±9.0%	11.0±9.3%	<0.001
Alanine aminotransferase (ALT)	74.2±19.4 IU/L	37.4±10.3 IU/L	<0.001

Data represented as mean±standard deviation unless otherwise specified. p-values calculated using independent Student t-test.

Table 3: Meta-analysis with pooled sensitivity, specificity and diagnostic odds ratio (DOR), and summary receiver operator characteristic (SROC) curve for detection of each level of fibrosis

	Studies (n _{total})	SROC	Sensitivity		Specificity		DOR	
			Pooled	Hetero- geneity	Pooled	Hetero- geneity	Pooled	Hetero- geneity
F1-4								
NFS (low threshold)	2 (432)	-	0.489 (0.432-0.546)	I ² =99.2%	0.816 (0.748-0.872)	I ² =97.4%	6.93 (3.84-12.53)	I ² =0.0%
NFS (high threshold)	2 (476)	-	0.766 (0.712-0.815)	I ² =0.0%	0.487 (0.406-0.569)	I ² =0.0%	3.10 (2.03-4.73)	I ² =0.0%
FIB4	2 (246)	-	0.398 (0.324-0.475)	I ² =98.0%	0.907 (0.817-0.962)	I ² =0.0%	3.75 (0.14-102.77)	I ² =92.4%
BARD	3 (504)	0.607 ±0.114	0.736 (0.668-0.796)	I ² =76.5%	0.362 (0.309-0.418)	I ² =81.3%	2.80 (1.44-5.43)	I ² =23.6%
MRE	2 (242)	-	0.514 (0.430-0.597)	I ² =63.9%	0.906 (0.830-0.956)	I ² =0.0%	10.51 (4.909-22.513)	I ² =0.0%
F2-4								
NFS (low threshold)	3 (471)	0.768 ±0.035	0.729 (0.652-0.797)	I ² =87.6%	0.509 (0.453-0.566)	I ² =96.2%	5.18 (2.91-9.23)	I ² =16.2%
NFS (high threshold)	2 (432)	-	0.351 (0.264-0.446)	I ² =90.5%	0.881 (0.840-0.914)	I ² =0.0%	2.26 (0.33-15.37)	I ² =80.6%
BARD	2 (343)	-	0.508 (0.416-0.601)	I ² =88.9%	0.574 (0.506-0.640)	I ² =96.5%	1.87 (1.16-3.03)	I ² =0.0%
FIB4	2 (343)	-	0.450 (0.359-0.543)	I ² =95.7%	0.919 (0.875-0.951)	I ² =93.9%	8.57 (4.59-15.97)	I ² =0.0%
Fibrotect (low threshold)	2 (509)	0.827 ±0.029	0.667 (0.590-0.737)	I ² =74.8%	0.751 (0.701-0.796)	I ² =0.0%	7.23 (3.02-17.29)	I ² =72.4%
Fibrotect (high threshold)	2 (555)	0.729 ±0.407	0.132 (0.070-0.219)	I ² =4.1%	0.989 (0.975-0.996)	I ² =36.3%	9.82 (3.28-29.39)	I ² =0.0%
ELF	2 (231)	-	0.709 (0.596-0.806)	I ² =0.0%	0.816 (0.745-0.874)	I ² =0.0%	9.71 (5.04-18.68)	I ² =0.0%
TE	8 (709)	0.851 ±0.017	0.758 (0.718-0.795)	I ² =85.5%	0.773 (0.730-0.813)	I ² =88.6%	13.90 (8.44-22.88)	I ² =39.9%
ARFI	3 (398)	0.837 ±0.024	0.740 (0.691-0.785)	I ² =94.1%	0.681 (0.622-0.737)	I ² =94.8%	10.41 (6.49-16.68)	I ² =0.0%
MRE	2 (242)	-	0.647 (0.522-0.759)	I ² =0.0%	0.937 (0.890-0.968)	I ² =22.7%	26.42 (11.19-62.38)	I ² =14.0%
F3-4								
NFS (low threshold)	17 (3388)	0.795 ±0.020	0.786 (0.755-0.814)	I ² =60.2%	0.622 (0.603-0.640)	I ² =94.4%	7.47 (5.47-10.20)	I ² =45.5%
NFS (high threshold)	13 (2733)	0.813 ±0.053	0.602 (0.567-0.637)	I ² =93.2%	0.924 (0.911-0.935)	I ² =94.4%	14.42 (7.81-26.61)	I ² =71.1%
BARD	13 (2804)	0.725 ±0.016	0.732 (0.697-0.765)	I ² =83.2%	0.578 (0.556-0.599)	I ² =88.3%	4.35 (3.52-5.37)	I ² =0.2%
FIB4 (low threshold)	10 (2202)	0.831 ±0.015	0.766 (0.729-0.800)	I ² =0.0%	0.726 (0.704-0.748)	I ² =86.0%	10.17 (7.33-14.09)	I ² =37.3%
FIB4 (high threshold)	7 (1392)	0.770 ±0.052	0.360 (0.307-0.417)	I ² =71.9%	0.950 (0.936-0.962)	I ² =92.6%	13.25 (8.77-20.03)	I ² =0.0%
Fibrometer	4 (827)	0.774 ±0.092	0.676 (0.621-0.728)	I ² =94.2%	0.691 (0.649-0.731)	I ² =87.0%	4.89 (1.71-14.04)	I ² =87.1%
Hepascore	2 (694)	-	0.693 (0.629-0.753)	I ² =21.0%	0.793 (0.754-0.829)	I ² =78.1%	9.83 (4.07-23.77)	I ² =77.2%
Fibrotect (low threshold)	2 (719)	0.359 ±0.244	0.832 (0.774-0.880)	I ² =38.2%	0.630 (0.587-0.672)	I ² =81.0%	12.02 (3.56-40.58)	I ² =62.1%
Fibrotect (high threshold)	2 (509)	0.847 ±0.062	0.461 (0.354-0.570)	I ² =82.1%	0.943 (0.916-0.963)	I ² =84.6%	14.03 (7.53-26.12)	I ² =0.0%
ELF	2 (248)	0.962 ±0.040	0.848 (0.750-0.919)	I ² =14.8%	0.916 (0.871-0.948)	I ² =31.8%	84.86 (20.23-356.0)	I ² =46.7%
TE	6 (1002)	0.859 ±0.017	0.827 (0.787-0.862)	I ² =89.0%	0.721 (0.689-0.752)	I ² =89.6%	16.65 (10.14-27.32)	I ² =32.5%
ARFI	3 (496)	0.902 ±0.035	0.789 (0.733-0.839)	I ² =91.5%	0.784 (0.745-0.820)	I ² =92.4%	24.25 (10.52-55.93)	I ² =62.4%
MRE	4 (486)	0.965 ±0.015	0.870 (0.792-0.927)	I ² =0.0%	0.915 (0.883-0.941)	I ² =0.0%	72.03 (35.62-145.62)	I ² =0.0%
F4								

TE	3 (386)	0.891 ±0.024	0.818 (0.722-0.892)	I ² =79.5%	0.783 (0.745-0.818)	I ² =94.2%	20.51 (10.69-39.38)	I ² =0.0%
ARFI	2 (361)	0.878 ±0.057	0.678 (0.569-0.774)	I ² =90.3%	0.818 (0.781-850)	I ² =95.7%	14.61 (5.22-40.95)	I ² =59.6%
MRE	2 (242)	-	0.789 (0.544-0.939)	I ² =5.7%	0.919 (0.875-0.951)	I ² =0.0%	41.26 (11.78-144.47)	I ² =0.0%

SROC – summary receiver operator characteristic curve; DOR – diagnostic odds ratio; NFS – NAFLD fibrosis

score, TE – transient elastography, ELF – enhanced liver fibrosis score; MRE – magnetic resonance

elastography.

Author Manuscript

Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:

Ooi, GJ;Mgaieth, S;Eslick, GD;Burton, PR;Kemp, WW;Roberts, SK;Brown, WA

Title:

Systematic review and meta-analysis: non-invasive detection of non-alcoholic fatty liver disease related fibrosis in the obese.

Date:

2018-02

Citation:

Ooi, G. J., Mgaieth, S., Eslick, G. D., Burton, P. R., Kemp, W. W., Roberts, S. K. & Brown, W. A. (2018). Systematic review and meta-analysis: non-invasive detection of non-alcoholic fatty liver disease related fibrosis in the obese.. *Obes Rev*, 19 (2), pp.281-294. <https://doi.org/10.1111/obr.12628>.

Persistent Link:

<http://hdl.handle.net/11343/261113>