Rhodes Freya A. (Orcid ID: 0000-0002-6884-6235)

SYSTEMATIC REVIEW: INVESTIGATING THE PROGNOSTIC PERFORMANCE OF FOUR NON-INVASIVE TESTS IN ALCOHOL-RELATED LIVER DISEASE.

Freya A Rhodes[1], Paul Trembling[1], Jasmina Panovska-Griffiths[2,3], Sudeep Tanwar[1,4], Rachel H Westbrook[1], Alison Rodger[3], William M Rosenberg[1]

[1]Institute for Liver and Digestive Health, UCL Division of Medicine, Royal Free Campus, London, UK
[2]Department of Applied Health Research, University College London, UK
[3]Institute for Global Health, University College London, UK
[4] Barts Health NHS Trust, London, United Kingdom

ORCID ID https://orcid.org/0000-0002-6884-6235

CORRESPONDENCE:

Professor William M Rosenberg: w.rosenberg@ucl.ac.uk
Address: Institute for Liver and Digestive Health, Division of Medicine,
University College London, Royal Free Campus, Rowland Hill Street, Hampstead,
London NW3 2PF, United Kingdom.
ORCID ID: 0000-0002-2732-2304

OTHER AUTHORS:

Dr Freya A Rhodes: f.rhodes@doctors.net.uk **ORCID ID** 0000-0002-6884-6235

Dr Sudeep Tanwar: sudeep.tanwar@nhs.net
Dr Paul Trembling: p.trembling@ucl.ac.uk
Professor Alison Rodger Alison.rodger@ucl.ac.uk
Dr Jasmina Panovska-Griffiths: j.panovska-griffiths@ucl.ac.uk
Dr Rachel Westbrook: rachel.westbrook@nhs.net

AUTHOR CONTRIBUTIONS

FR designed the study, created the search strategy, conducted the literature search and data extraction, analysed the results and wrote the first draft of the manuscript. PT acted as second reviewer, reviewing the search results and extracting data independently. J PG, ST, RW and AR reviewed and edited the manuscript. WR conceived the study, participated in the design

of the study and edited drafts of the manuscript. All authors read and approved the final manuscript.

ETHICS AND PATIENT CONSENT

Ethical approval was not required for this systematic review, since it used data from previous studies which had their own ethics and patient consent.

Electronic word count (main text): 4,984

Number of tables = 4

Number of figures = 3

ABBREVIATIONS:

FIB 4: Fibrosis 4 test

ELF: Enhanced Liver Fibrosis test

HR: Hazard Ratio

RR: Relative Risk

CI: Confidence Interval

AUROC: Area Under the Receiver Operating Characteristics curve

ArLD: Alcohol related Liver Disease

NIT: Non-Invasive Test

US: Ultrasound

NAFLD: Non-alcoholic Fatty Liver Disease

AFLD: Alcoholic Fatty Liver Disease

APRI: AST to Platelet Ratio Index

CLD: Chronic Liver Disease

LRE: Liver Related Event

AST: Aspartate transaminase

ALT: Alanine aminotransferase

G/day = grams per day

LRM = Liver Related Mortality

ACM= All-cause Mortality

LFTs = liver function tests

PRISMA = Preferred Reporting Items for Systematic reviews and Meta-Analyses

PICO = Participants, interventions, comparators, outcomes and study design

QUIPS = Quality In Prognosis Studies

MeSH = Medical Subject Headings

WOS: Web of Science

ALP: Alkaline phosphatase

IQR: Inter Quartile Range

MELD: Model for End stage Liver Disease

CP: Child Pugh score

SD: Standard deviation

HCC: Hepatocellular Carcinoma

HE: Hepatic Encephalopathy

BMI: Body Mass Index

HTN: Hypertension

SNLRD = Survival or Non-liver related Death

LSM = Liver Stiffness Measurement

COMPETING INTERESTS STATEMENT:

ABSTRACT:

Background/Aims:

Mortality of Alcohol-related-Liver-Disease (ArLD) is increasing, and liver fibrosis stage is the best mortality predictor. Non-invasive-tests (NIT) are increasingly used to detect fibrosis, but their value as prognostic tests in chronic liver disease (CLD), and in particular in ArLD is less well recognized. We aimed to describe the prognostic performance of four widely used NITs (FIB4, ELF test, FibroScan and FibroTest) in ArLD.

Methods:

Applying systematic-review methodology, four databases were searched from inception to May 2020. Inclusion/exclusion criteria were applied to search using MeSH terms and keywords. First and second reviewers independently screened results, extracted data and performed risk-of-bias assessment using Quality-In-Prognostic-Studies (QUIPS) tool.

Results:

Searches produced 25,088 articles. After initial screening, 1,020 articles were reviewed independently by both reviewers. Eleven articles remained after screening for eligibility: one on ELF, four on FibroScan, four on FIB4, one on FIB4+FibroScan and one on FibroTest+FIB4. Area-Under-Receiving-Operator-Characteristics-curves (AUROCS) for outcome-prediction ranged from: 0.65-0.76 for FibroScan, 0.64-0.83 for FIB4, 0.69-0.79 for FibroTest and 0.72-0.85 for ELF. Studies scored low-moderate risk of bias for most domains, but high-risk in confounding/statistical reporting domains. The results were heterogeneous for outcomes and reporting, making pooling of data unfeasible.

Conclusions:

This systematic-review returned eleven papers, six of which were conference-abstracts and one unpublished manuscript. Whilst the heterogeneity of studies precluded direct comparisons of NITs, each NIT performed well in individual studies in predicting prognosis in ArLD (AUROCs >0.7 in each NIT category), and may add value to prognostication in clinical practice.

KEYWORDS: *(Liver cirrhosis), (liver diseases, alcoholic), (non-invasive test), (Alcohol Use Disorder) (FIB4), (FibroTest), (ELF), (FibroScan), (prognosis), (mortality).*

KEY POINTS

- Few studies focus on NIT performance in alcohol compared to other aetiologies

- Overall, we found good performance (AUROC >0.7) for FIB4, FibroScan, FibroTest and ELF in predicting mortality or liver-related-events.

- In the single study on ELF, ELF outperformed histology in predicting prognosis in ArLD, in keeping with its performance in all-cause CLD (1).

- In the single biopsy-paired study on FibroTest, Fibrotest performed at least as well as histology in predicting outcomes.

- FIB4 appears to perform better than MELD in prognosis prediction in ArLD.

- FIB4/FibroScan/FibroTest/ELF can be practical tools to aid prognostication in ArLD

INTRODUCTION

Mortality rates from cirrhosis have increased by 400% over the last 30 years, largely attributed to alcohol (2). The degree of liver fibrosis is the strongest predictor of mortality in chronic liver disease,(3) and thus it is important for clinicians to have information about fibrosis in order to predict clinical outcomes and guide individualised treatment decisions. Liver biopsy is the traditional modality for detecting and quantifying fibrosis in alcohol-related liver disease (ArLD) and the current reference standard against which other tests for fibrosis are evaluated. However liver biopsy is considered an imperfect test owing to its invasive nature with associated risks to the patient, as well as sampling error and reporting bias (4, 5). Therefore, there has been a drive to develop non-invasive tests (NIT) for liver fibrosis over the last two decades to assess fibrosis severity and to determine prognosis. These NIT largely comprise blood tests that measure direct and indirect markers of liver fibrosis, of which the most widely studied are The Enhanced Liver Fibrosis (ELF) test, FibroTest, HepaScore, Fibrometer, FIB4, Forns' Index, APRI, AST:ALT ratio, and age-platelet index (6). There are also "physical" techniques assessing liver stiffness, including FibroScan, sheer-wave elastography and MR elastography but these are less generalisable due to operator performance and availability.

A scoping exercise was conducted to identify NIT that had been investigated for both their prognostic and diagnostic performance, and were established enough that they could be readily translated into clinical practice for routine prognostic assessment. The markers selected on these criteria are: FibroScan, FIB4, ELF test and FibroTest. The selection of prognostic markers is of particular importance in the practice of stratified or personalised medicine where they can support clinicians and patients in making decisions about management such as initiating treatments, and initiating enhanced monitoring for complications of cirrhosis.

Whilst there is an increasing number of studies on prognostic markers, few have been externally validated for use in clinical practice (7). Moreover, the majority of validation studies have been performed in patients with either viral hepatitis or unselected chronic liver disease, rather than specifically in ArLD. It has been shown in cholangiopathies and all-cause CLD that NIT can out-perform histology in predicting clinical outcomes (8, 9), and therefore it is of great clinical importance to know if NIT can also reliably predict outcomes in ArLD, the commonest aetiology of cirrhosis.

This systematic review aims to determine the prognostic performance of four commonly used NIT for liver fibrosis in ArLD, specifically in predicting mortality, and liver related events (LRE) resulting in decompensated cirrhosis and death.

PATIENTS AND METHODS

This systematic review was conducted using the guidance described in the Cochrane Handbook of Interventions (10). The aim of this study was to identify the prognostic performance of four non-invasive tests for liver fibrosis in alcohol-related liver disease – FibroScan, ELF test, FibroTest and FIB4. The PICO structure (participants, interventions, comparators, outcomes and study design) was used and PRISMA guidance followed (Preferred Reporting Items for Systematic reviews and Meta-Analyses) (See supplementary table 1). The protocol for this review was registered with PROSPERO (CRD42020175605).

PICO

Participants: All adult humans (age 18+ years) with Alcohol related Liver Disease

Intervention: Studies that included FIB4, FibroTest, FibroScan or ELF test as prognostic markers in ArLD were included.

Comparisons: Each of the above interventions were compared to one another.

Outcomes:

1. The ability of ELF, FibroTest, FibroScan and FIB4 to predict all cause and liver-related mortality
2. The ability of ELF, FibroTest, FibroScan and FIB4 to predict liver-related cirrhotic decompensation events including ascites, variceal bleeding, encephalopathy, need for liver transplantation and development of hepatocellular carcinoma (HCC)

SEARCH STRATEGY

Four databases were searched systematically, using search strategies which can be found in supplementary tables 2, 3 and 4.

Search themes related to the study PICO, including a combination of MeSH terms and keywords. Pilot searches were conducted in order to refine the search strategy.

Firstly, Web of Science, Ovid Medline, Embase and Cochrane Library were systematically searched (see supplementary table 1). Secondly, reference lists of included studies and

relevant review articles were hand-searched to identify any further potentially relevant publications. Thirdly, where information from abstracts or full texts was not sufficient for us to include the study, we contacted relevant authors by email to request data.

References were imported into EndnoteTM web basic reference manager, and then the selection of articles for both authors to review was imported into Rayyan systematic review manager (11), which enabled independent, blinded review of each article and documentation of reasons for exclusion.

The first reviewer (FR) compiled the search strategy, performed the first search, and conducted the first sift of journal articles by title and abstract. The abstracts of the remaining 1,020 articles were then reviewed by the first and second reviewer (PT) independently using Rayyan systematic review manager (11). Articles were selected using the pre-defined inclusion and exclusion criteria (table 1). Reasons for exclusion were documented, and where there was any discrepancy in decision, this was able to be resolved by discussion between the two reviewers, or by input from third reviewer (WR) when consensus not achieved. The resulting articles were then reviewed by full text independently by the first and second reviewer, and a final list of articles for inclusion was created (See figure 1).

SELECTION CRITERIA

See table 1 for the full list of inclusion and exclusion criteria. All levels of evidence were included apart from descriptive review articles and opinion pieces. Non-human and pre-clinical studies were excluded. No restriction was made on language. Grey literature (conference abstracts and unpublished work) was not excluded, in line with Cochrane guidance (10). Due to the paucity of prognostic biomarker data on ArLD, relevant studies on chronic liver disease in general were included, as long as they incorporated at least 10 patients where the primary aetiology was alcohol, and that these alcohol data could be extracted separately. Studies were required to have reported either relative risk (RR), hazard ratio (HR) or Area Under the Receiver Operating Characteristic curve (AUROC), with corresponding confidence intervals (CI), for data extraction in order to address prognosis.

DATA EXTRACTION STRATEGY

Data extraction was undertaken by both reviewers independently (FR and PT) using a pre-defined data-entry form (see supplementary figure 1). Any disagreements were resolved through discussion between FR and PT, or with WR if persisting uncertainty.

Information collected included journal title, year, type of publication, type of non-invasive test investigated, number of patients in study cohort, number of patients with alcohol as primary aetiology, patient demographics, alcohol consumption data, statistical methods used and test performance characteristics (See tables 2 and 3). Where studies had included a comparison of non-invasive test with histology, this was recorded and evaluated.

If more than one publication included data from the same cohort of patients, the data from the most recent and comprehensive report were included to avoid duplication in line with Cochrane methodology.

Where data were not clear, or where data were reported for chronic liver disease in general but not specifically for those patients with alcohol as the aetiology, authors were contacted by email to request clarification or access to their original data. If the author failed to reply, the study was excluded.

QUALITY ASSESSMENT

The quality of the included prognostic studies was assessed by two authors (FR and PT) independently using the QUIPS (Quality In Prognosis Studies) tool (12).

This allowed grading of each publication for risk of bias as being at low, medium or high based on six domains: study participation, study attrition, prognostic factor measurement, outcome measurement, study confounding and statistical analysis and reporting. Disagreements between the assessors were resolved through discussion. This process identified the need to clarify the prognostic factor measurement. Originally a 'low' risk of bias was assigned only to those studies in which the NIT failed in less than 10%. After discussion it was recognised that blood biomarkers have a result reporting success rate of 100% and this failure rate only applied to studies using FibroScan. The QUIPS tool can be found in supplementary table 6.

DATA SYNTHESIS AND ANALYSIS

Prognostic outcomes were reported as HR or AUROC with 95% confidence intervals. Where RR were reported, these were taken to be equivalent to HR. Due to the heterogeneity and small number of final included studies, a descriptive approach was taken to analyse the results.

RESULTS

Study Selection

Searching the four databases returned 25,088 results, of which 8,781 were duplicates (detected and removed using Endnote). An additional 8 results were found by searching reference lists of included papers and relevant review articles. Three of the results (two conference abstracts (13, 14) and one full paper (1)) reported data from patients with ArLD from the same patient cohort. The full paper (1) did not detail the prognostic performance of ELF in the ArLD cohort separately from the mixed-aetiology liver patients, and so was excluded. The senior author of the most recent abstract (13) was contacted, and permission gained to use unpublished data from this abstract that had been written up as a manuscript under review for publication (15). As this was the most recent and comprehensive of the articles, this was included even though it is not yet published, and the older conference abstract (14) reporting the same cohort was excluded.

We found several articles investigating the prognostic performance of non-invasive tests in mixed aetiology chronic liver disease that did not separately specify the performance in

ArLD patients, and therefore these studies were excluded from this review (16-38). One systematic review (39) was excluded as it was also investigating prognostic performance of FibroScan in mixed aetiology liver disease – and although it referenced studies which included patients with ArLD (38, 40-43), these studies either had unrelated outcomes, the sample size of the alcohol patients was less than n=10, or they did not report data on patients with ArLD separately.

This resulted in 16,316 articles that were screened by title or abstract by the first author, with 15,296 being excluded, leaving 1,020 publications for review of abstract by the first and second reviewer independently. The full texts of the 40 selected articles were assessed for eligibility by the first and second reviewer, and, after resolving discrepancies, 11 articles remained for inclusion in the data analysis. This comprised the unpublished manuscript, 4 full-text published papers and a further 6 conference abstracts (see figure 1). We found no systematic reviews that specifically reported the prognostic performance of any of the desired four non-invasive tests in ArLD.

STUDY CHARACTERISTICS

Of the eleven studies included, four evaluated FIB4 (three full papers, one abstract), four FibroScan (all abstracts), one evaluated both FIB4 and FibroScan (abstract), one study evaluated both FIB4 and FibroTest (full paper), and one study evaluated ELF (full paper, unpublished). The total number of patients with ArLD included in the analyses of these eleven studies was 20,412, with a median number of participants of 218 (range 64 - 17,300). Seven studies were prospective, two were retrospective and two were unspecified. The general characteristics for each study, with references, are detailed in tables 3 and 4. Studies were conducted between 2009 and 2019. The median age of participants was 48.5 (range 41.6-60), and 74% were male (range 62.9-100%). The median background prevalence of cirrhosis or advanced fibrosis was heterogeneous, with one study excluding patients with known cirrhosis from the outset (recruiting patients with fatty liver on imaging and a significant alcohol history), six reporting 100% with cirrhosis, one with 31% cirrhosis (biopsy-proven) and one with 77.8% cirrhosis (biopsy proven), and a further two did not specify.

Outcomes were also heterogeneous, and included liver-related events (LRE), development of HCC, index variceal bleed, liver-related mortality (LRM) and all-cause mortality (ACM). Eight of the eleven articles exclusively investigated patients with alcohol-related liver disease, and three investigated people with chronic liver disease of mixed aetiology, but included details of sub-group analyses, specifying results for patients with alcohol-related liver disease within their cohorts.

The significant heterogeneity of these studies precluded meta-analysis or pooling of results.

RISK OF BIAS WITHIN STUDIES

On review of the six bias domains in the QUIPS tool, the majority of the eleven included studies were assessed to be at low or moderate risk of bias in study attrition, prognostic factor measurement and outcome measurement. However, there were some studies that were at high risk of bias in the other three domains. In the first domain (study participation), 4/11 studies scored 'high risk'. Three of these were conference abstracts, and one was the unpublished manuscript which incorporated pooled data from three patient cohorts, but did not clearly specify the time period for each of the studies.

In the study confounding domain, 6/11 (55%) of the studies scored 'high risk of bias', due either to not defining the confounding variables, adjusting for fewer than three confounding variables, or reporting an unadjusted analysis. Four out of six of these articles were conference abstracts and so may have omitted this information because of restrictions on word count. The other two (13, 44) had either only adjusted the analysis for two variables or no adjustment was documented, and thus were both graded as being at high risk of bias using the QUIPS tool. In the final domain 'statistical analysis and reporting', three of the eleven articles were graded as being at high risk because they either did not report multivariable analysis, or they reported HRs or AUROCs without corresponding confidence intervals. These three were all conference abstracts.

Overall, 78.8% of the six domains across ten studies were rated 'Low' or 'Moderate' risk of bias (See figure 2 and supplementary table 7). Cohen's kappa (κ) was measured to determine if there was agreement between the two reviewers on the grading of low, moderate and high risk of bias across six domains over the 11 articles. This showed moderate-to-good agreement (45), with 74.2% of all grades (6 domains x 11 papers) being the same between the two reviewers, and $\kappa = 0.59$ (95%CI, 0.426 to 0.75), p = <0.0005.

PROGNOSTIC PERFORMANCE OF EACH OF THE FOUR NON-INVASIVE TESTS
(See table 4, figure 3)

FIB4

Six studies examined the prognostic performance of FIB4 in ArLD, three of which used Liver Related Mortality (LRM) as the outcome. The fourth study used 'development of HCC at 3 years' as the outcome, the fifth used index variceal bleed within 6 months, and the sixth used 'mortality' – which did not specify if liver-related or all-cause.

AUROCS for mortality were recorded in 3 studies, and Hazard ratio (HR) in the other mortality study. AUROCS were 0.64 (95%CI 0.55-0.71), 0.78 (no CI reported) and 0.825 (95%CI 0.71-0.93). The study which reported HR for LRM at 14 years showed a significant difference in mortality based on FIB4 thresholds, with the higher threshold of >2.67 giving a HR of 32.9 (95%CI 15.04-71.96) compared with a low FIB4 threshold of <1.3 (HR 1.14, 95% CI 0.34-3.85).

Four studies used continuous FIB4 score in their analysis, and two used FIB4 thresholds, which were different in each study. One study identified three categories of FIB4 score: low (1.3), intermediate (1.3-2.67) and high >2.67 and the other used a single threshold categorising results above or below 3.23. The latter study used development of HCC at 3 years as the main outcome, and FIB4 was able to predict this with AUROC of 0.69 (0.63-0.75).

Two studies examined FIB4 along with another non-invasive test (FibroTest in one and FibroScan in the other). Cho et al. (46) found no reported difference between AUROC for FIB4 (AUROC 0.78) and FibroScan (AUROC 0.73) in predicting LRM (although p values were not stated for this comparison). However, when using a multivariable cox proportional hazard model, FIB4 was able to predict LRM (HR 1.11, p= 0.03) but FibroScan was not. FIB4 was also able to independently predict ACM (p = <0.001) whereas FibroScan was not (confidence intervals were not reported).

Naveau et al. (47) compared FIB4 with FibroTest. Although there was a statistically significant difference between the AUROCs for the tests for prediction of liver related death (FibroTest AUROC 0.79 (95%CI 0.68-0.86), FIB4 AUROC 0.65 (95% CI 0.54-0.74); p=0.004) there was no significant difference in AUROCS for predicting overall survival (FibroTest AUROC 0.69 (95%CI 0.61-0.76), FIB4 AUROC 0.64 (0.55-0.71); p = 0.20).

FibroScan

Five studies investigated the prognostic performance of FibroScan. One of them (46) as described above, compared FibroScan (continuous liver stiffness measurement) with FIB4, and found that FIB4 was able to predict mortality in multivariable cox proportional hazard analysis, but FibroScan was not, although AUROCS were not significantly different between FibroScan (0.73) and FIB4 (0.78).

Four other studies (48-51) reported AUROCs for predicting mortality or liver-related events using continuous liver stiffness measurements, with AUROCs of 0.65, 0.675, 0.7 and 0.76. Two of these four studies (48, 51) also reported liver stiffness thresholds, with one using a threshold of 25kPa, finding a significant difference in mean incidence of LRE of 4.5% in the <25kPa cohort compared with 15.5% in the >25kpa cohort (51). The other study (48) reported a liver stiffness threshold of 20kPa, with incidence of death at 3% in patients with liver stiffness measurement (LSM) <20kPa, compared to 15% deaths in those with LSM >20kPa (p = <0.004).

FibroTest

Only one study reported the prognostic performance of FibroTest in ArLD (47). This study of 218 people with ArLD compared FibroTest with liver biopsy, hepascore, fibrometer, FIB4, APRI and Forns' Index in predicting LRM and ACM. FibroTest performed better than FIB4 (p= 0.004) in predicting LRM (FibroTest AUROC 0.79 (95% CI 0.68-0.86) compared to FIB4 AUROC of 0.65 (95% CI 0.54-0.74) (although there was no difference between FIB4 and FibroTest in predicting ACM). When compared with other markers of fibrosis in this study, the prognostic values of FibroTest (AUROC 0.79 ± 0.04 for survival or non-liver disease related death), Hepascore (0.78 ± 0.04), and fibrometer (0.80 ± 0.04) did not differ from that of liver biopsy fibrosis staging (0.77 ± 0.04). In multivariate analysis, they found that the best performing tests were FibroTest (p = 0.004) and liver biopsy (p = 0.03).

ELF

We found only one study (manuscript under review) (15) that investigated the prognostic performance of the ELF test in ArLD. Data from this study have been published by the same authors as a conference abstract (13).

This study comprised 64 people with ArLD taken from three different study cohorts with a total sample size (of mixed aetiology liver disease) of 786.

ELF was analysed both as a continuous value and using thresholds of 9.8, 10.5 and 11.3. When analysed as a continuous value, ELF predicted LRE at 6, 7 and 8 years with AUROCs of 0.816, 0.844 and 0.847 respectively. Risk ratio for the prediction of LRE at 6 years was 1.82 (1.169-2.83). ELF also predicted All-Cause Mortality (ACM) at 6,7 and 8 years with AUROCs of 0.733, 0.722 and 0.722 respectively. When analysed using a cox proportional hazard model (adjusted for age and gender), Connoley et al. found that each unit increase in ELF was associated with a 1.44 times increased risk of LRE (95% CI 1.25-1.66, p <0.001). When analysed, the HR for ELF scores between 10.5-11.29 was 3.84 (95% CI 0.90-16.39), HR for ELF ≥11.3 was 10.24 (95% CI 2.97-35.27), compared to a low ELF threshold of <9.8 where HR was 1.49 (95% CI 0.287-7.74).

Performance of NIT compared with histology and other prognostic scores:

Only 2 out of 11 studies were biopsy-paired (15, 47). Whilst Naveau et al found that FibroTest performed equally as well as histology, (47) Connoley et al found that ELF was superior to histology in predicting prognosis in ArLD (ELF AUROC for all-cause mortality 0.733 (95% CI:0.645-0.861), compared to 0.600 (95% CI:0.470-0.730) for liver biopsy, p = <0.05) (15).

Where one of the four NITs of interest were directly compared with other more traditional prognostic scores such as MELD and Child Pugh (CTP), FibroTest outperformed CTP (FibroTest AUROC for survival 0.79 (95%CI 0.68-0.86), CTP AUROC 0.69 (95%CI 0.58-0.77), p=0.02) (47), and FIB4 outperformed MELD in two separate studies: In Chaudhari et al's study, FIB4 AUROC for mortality was 0.83 (95%CI 0.71-0.93), MELD 0.70 (0.53-0.87) p=0.001 (52), and in Kothari et al's study, FIB4 AUROC for predicting variceal bleed was 0.74 (95%CI 0.66-0.81), MELD AUROC 0.54 (95%CI 0.46-0.62) (53).

Whilst MELD and CTP are very much still a part of current clinical practice for prognosticating in CLD, these findings suggest that non-invasive fibrosis markers may be a better choice in predicting prognosis in ArLD.

DISCUSSION

Main findings

Whilst there is now good evidence for the use of NIT for prognosticating in chronic liver disease of mixed aetiologies (16, 18-38, 54), this systematic review has found fewer studies on ArLD. This is important, as each aetiology of liver disease behaves differently in terms of pathophysiology, clinical presentation and complications and there is evidence suggesting that the performance of some NIT varies with disease aetiology (55-60). Whilst non-invasive tests are becoming increasingly widely used both for the diagnostic staging of liver fibrosis as well as prognosis, it is imperative that they are evaluated in representative populations with specific aetiologies.

Mortality rates for cirrhosis have increased by 400% in the last thirty years, with alcohol-related liver disease identified as the predominant cause (2). There is thus an international growing recognition of the importance of research in alcohol-related liver disease, and this study has highlighted the gaps in current knowledge of commonly used non-invasive liver fibrosis tests as prognostic markers in this condition.

Nevertheless, all the four NIT investigated in this review show promising prognostic performance, with AUROCS above 0.7 in some studies for each test, and AUROCS above 0.8 for one study of FIB4 (52) and the single ELF study (15). While heterogeneity of the results prevented statistical comparisons of test performance between studies there were two studies which directly compared two NITs. The study which compared FibroTest to FIB4 (47) found a significantly better prognostic performance of FibroTest when compared to FIB4. The other study which directly compared FIB4 with FibroScan (46), found that FIB4 could predict LRD (HR 1.11; p=0.003) while FibroScan could not.

However, it is not possible to identify a single NIT that performs better than the rest based on this systematic review due to the heterogeneity and lack of data.

The ELF test did appear to be superior to histology in predicting outcomes in one study in ArLD (15), and FibroTest performed equally well to histology (47). FibroTest also outperformed CTP (47) and FIB4 outperformed MELD in two studies in ArLD (52, 53). Biopsy is not routinely required in the management of ArLD, and the findings from this review indicate that these commonly available NIT can perform a useful role when predicting prognosis in clinical practice for patients with ArLD, avoiding the need for liver biopsy and may even be superior to more traditional prognostic scores like MELD and CTP.

Strengths

The strengths of this study are based on a comprehensive search, with rigorous screening of texts. The screening of texts, extraction of data, and assessment of risk of bias were all conducted by the first and second reviewers independently, to minimise reporting bias. Where it was not possible to extract data from certain papers, efforts were made to contact authors by email. In addition, reference lists of relevant articles were hand-searched to maximise inclusion for study, and grey literature was included. Where data were found to be using the same patient cohort – only the most recent and comprehensive study was used, to prevent risk of duplication of results.

Limitations

We recognise there are limitations of this study. The heterogeneity and low number of included studies prevented meta-analysis, or pooling of hazard ratios or AUROCS.

For example, of the four studies on FIB4 that all reported AUROCs for mortality, one reported LRM, and the three that reported ACM had different end points (between 1 and 5 years), and two out of these three did not report corresponding confidence intervals. This variety in outcomes and reporting made it impossible to pool results, and is in keeping with existing wider literature on prognostic studies in all fields that recognises the variable quality of prognostic studies. The Cochrane methods group has acknowledged the challenges in systematic reviews on prognosis due to "low quality of primary studies, poor reporting, and difficulties in combining results across different research designs, analyses and presentations of results"(61). Other studies have commented on the barriers to synthesis of prognostic study data ranging from poor reporting, lack of consistency in statistical analysis across prognostic studies and often prognostic model studies are based on relatively small sample sizes leading to overfitting and poor generalisability of results (62, 63). A clear finding arising from the conduct of this review is the necessity for larger rigorous studies of NIT in ArLD.

Six out of eleven of these studies were conference abstracts, limiting the data that could be extracted, leading to higher scores on the risk of bias assessment.

As all of the studies reported significant results for the performance of non-invasive tests, it is possible there was publication bias. However, it is difficult to deduce this for certain as the number of publications in this area is so small. The inclusion of grey literature in this systematic review may have reduced publication bias.

Conclusions

This study has demonstrated that all four of the examined NIT can perform well in predicting prognosis in ArLD with AUROCs >0.7. Of the two included studies that were biopsy-paired, they found that FibroTest and ELF performed as well and better than histology in predicting outcomes, respectively. Whilst the heterogeneity of studies precluded pooling of results and direct comparisons, those studies that did include direct comparisons of NIT with other 'traditional' prognostic scores showed NIT for liver fibrosis to be superior to MELD and CTP. With easy availability of FIB4/FibroTest/Fibroscan/ELF, it therefore may be preferable to use one of these fibrosis markers when prognosticating in ArLD in situations where biopsy is not necessary. Due to the small number of included studies, further validation studies of these NIT as prognostic scores are warranted.

REFERENCES

1.   Parkes J, Roderick P, Harris S, Day C, Mutimer D, Collier J, et al. Enhanced liver fibrosis test can predict clinical outcomes in patients with chronic liver disease. Gut. 2010;59(9):1245-51.

2.   Robyn Burton CH, Don Lavoie et al. The Public Health Burden of Alcohol and the Effectiveness and Cost-Effectiveness of Alcohol Control Policies: An evidence review [webpage]. https://www.gov.uk/government/publications/the-public-health-burden-of-alcohol-evidence-review: Public Health England; 2016 [updated 02.12.2016; cited 2016 2016]. Public Health England document]. Available from: https://www.gov.uk/government/publications/the-public-health-burden-of-alcohol-evidence-review.

3.   Lackner C, Spindelboeck W, Haybaeck J, Douschan P, Rainer F, Terracciano L, et al. Histological parameters and alcohol abstinence determine long-term prognosis in patients with alcoholic liver disease. J Hepatol. 2017;66(3):610-8.

4.   Maharaj B, Maharaj RJ, Leary WP, Cooppan RM, Naran AD, Pirie D, et al. Sampling variability and its influence on the diagnostic yield of percutaneous needle biopsy of the liver. Lancet. 1986;1(8480):523-5.

5.   Regev A, Berho M, Jeffers LJ, Milikowski C, Molina EG, Pyrsopoulos NT, et al. Sampling error and intraobserver variation in liver biopsy in patients with chronic HCV infection. Am J Gastroenterol. 2002;97(10):2614-8.

6.   Sharma S, Khalili K, Nguyen GC. Non-invasive diagnosis of advanced fibrosis and cirrhosis. World Journal of Gastroenterology. 2014;20(45):16820-30.

7.   Altman DG, Vergouwe Y, Royston P, Moons KG. Prognosis and prognostic research: validating a prognostic model. BMJ. 2009;338:b605.

8.   Parkes J, Roderick P, Harris S, Day C, Mutimer D, Collier J, et al. Enhanced liver fibrosis test can predict clinical outcomes in patients with chronic liver disease. Gut. 2010;59(9):1245-51.

9.   Mayo MJ, Parkes J, Adams-Huet B, Combes B, Mills AS, Markin RS, et al. Prediction of clinical outcomes in primary biliary cirrhosis by serum enhanced liver fibrosis assay. Hepatology. 2008;48(5):1549-57.

10.   Higgins J, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al. Cochrane Handbook for Systematic Reviews of Interventions version 6.0 Cochrane2019 [updated July 201917/03/2020]. Available from: www.training.cochrane.org/handbook.

11.   Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan-a web and mobile app for systematic reviews. Syst Rev. 2016;5(1):210.

12.   Hayden JA, van der Windt DA, Cartwright JL, Cote P, Bombardier C. Assessing bias in studies of prognostic factors. Ann Intern Med. 2013;158(4):280-6.

13.   Connoley D, Patel P, Hogan B, Tanwar S, Rhodes F, Parkes J, et al. The utility of the enhanced liver fibrosis test in alcoholic liver disease. J Hepatol. 2019;70 (1):e815.

14.   Trembling PM, Parkes J, Tanwar S, Burt AD, Rosenberg WM, Investigators ELF. ENHANCED LIVER FIBROSIS TEST ACCURATELY IDENTIFIES LIVER FIBROSIS AND PREDICTS CLINICAL OUTCOMES IN ALCOHOLIC LIVER DISEASE. J Hepatol. 2012;56:S424-S.

15.   Connoley D, Patel P, Hogan B, Tanwar S, Rhodes F, Parkes J, et al. The Enhanced Liver Fibrosis test maintains its diagnostic and prognostic performance in alcohol-related liver disease. . (Manuscript in preparation). 2020.

16.   Hashim A, Haddadin Y, Macken L, Bremner S, Keller M, File A, et al. Can transient elastography predict development of portal hypertension and or hepatic decompensation in individuals with cirrhosis? Gut. 2017;66 (Supplement 2):A49.

17.   Cho E, Lee JG, Sohn JH, Jeong JY, Kim TY, Kim SM, et al. Combined Effect of hepatic venous pressure gradient and liver stiffness on long-term mortality in patients with cirrhosis. Journal of Gastroenterology and Hepatology. 2018;33:401-.

18.   Hashim A, Parnell B, Haddadin Y, Macken L, Bremner S, Keller ME, et al. Does liver stiffness measurement and controlled attenuation parameter predict portal hypertension, liver related events and overall mortality in individuals with cirrhosis? Hepatology. 2017;66 (Supplement 1):263A-4A.

19.   Day JW, Rosenberg WM, Parkes J. ELF test thresholds for disease stratification and prognosis in chronic liver disease. Hepatology. 2015;62:596A.

20.   Hogan BJ, O'Beirne J, Patch DW, Yu D, Parisi I, Dhillon AP, et al. ELFTM predicts clinical outcomes in liver disease. Hepatology. 2015;62:604A-5A.

21.   Stauber RE, Spindelboeck W, Putz-Bankuti C, Pock H, Stojakovic T, Obermayer-Pietsch B. Enhanced Liver Fibrosis (ELF) score predicts mortality in cirrhosis. Hepatology International. 2013;7:S492-S3.

22.   Parkes J, Roderick P, Harris S, Gough C, Wheatley M, Alexander GJ, et al. European liver fibrosis (ELF) panel of serum markers can predict clinical outcome in a cohort of patients from England with mixed aetiology chronic liver disease. Hepatology. 2007;46(4):832A-A.

23.   Adler M, Larocca L, Trovato FM, Marcinkowski H, Pasha Y, Taylor-Robinson SD. Evaluating the risk of hepatocellular carcinoma in patients with prominently elevated liver stiffness measurements by FibroScan: a multicentre study. HPB. 2016;18(8):678-83.

24.   Singh S, Fujii LL, Murad MH, Wang Z, Asrani S, Ehman R, et al. Liver stiffness measurement predicts risk of decompensation, hepatocellular cancer and mortality in patients with chronic liver diseases: A systematic review and meta-analysis. Hepatology. 2013;58(4):956A.

25.   Munteanu M, Rudler M, Lebray P, Perazzo H, Massard J, Pais R, et al. Long term prognostic value of the FibroTest in patients with non-alcoholic-fatty-liver disease (NAFLD), compared to chronic hepatitis C (CHC), B (CHB), and alcoholic liver disease (ALD). J Hepatol. 2018;68 (Supplement 1):S99-S100.

26.   Boursier J, Bertrais S, Marsault P, Rapilly P, Fouchard-Hubert I, Oberti F, et al. Prediction of mortality by fibroscan, Fibrometer and their combination in a cohort of 3,623 patients with chronic liver disease. Hepatology. 2011;54:1227A.

27.   Ngo Y, Perazzo H, Munteanu M, Lebray P, Moussalli J, Thabut D, et al. Meta-Analysis of the prognostic value of liver fibrosis biomarkers. Hepatology. 2011;54:1225A.

28.   Rouyer M, Royer L, Perazzo H, Munteanu M, Ngo Y, Luckina E, et al. Non-invasive prediction of upper gastrointestinal bleeding by rupture of esophageal varices (ROV) using liver and spleen stiffness (Aixplorer), liver stiffness by fibroscan and fibrotest. J Hepatol. 2013;58:S96.

29.   Irvine KM, Wockner LF, Shanker M, Fagan KJ, Horsfall LU, Fletcher LM, et al. Predicting clinical outcomes in chronic liver disease: The ELF test is superior to histology and simple scores. Journal of Gastroenterology and Hepatology (Australia). 2015;30:104-5.

30.   Kim JH, Choe WH, Kwon SY. Predictive value of MELD, APRI, and FIB-4 for mortality and development of HCC in cirrhotic patients. Hepatology International. 2016;10(1):S356.

31.   Buechter M, Kahraman A, Manka P, Gerken G, Jochum C, Canbay A, et al. Spleen and Liver Stiffness Is Positively Correlated with the Risk of Esophageal Variceal Bleeding. Digestion. 2016;94(3):138-44.

32.   Festi D, Colecchia A, Schiumerini R, Marzi L, Mandolesi D, Reggiani MLB. Spleen and liver stiffness measurement can predict clinical complications in compensated cirrhotic patients: A prospective study. Hepatology. 2012;56:926A.

33.   Kumar A, Shastri SM, Kumar M, Sharma P, Garg H, Sarin SK. Transient elastography correlates with portal pressure and identifies variceal bleeders. Indian Journal of Gastroenterology. 2010;29(1):A96-A7.

34.   Sporea I, Ratiu I, Sirli R, Popescu A, Bota S. Value of transient elastography for the prediction of variceal bleeding. World Journal of Gastroenterology. 2011;17(17):2206-10.

35.   Klibansky DA, Mehta SH, Curry M, Nasser I, Challies T, Afdhal NH. Transient elastography for predicting clinical outcomes in patients with chronic liver disease. Journal of Viral Hepatitis. 2012;19(2):e184-e93.

36.   Robic MA, Procopet B, Metivier S, Peron JM, Selves J, Vinel JP, et al. Liver stiffness accurately predicts portal hypertension related complications in patients with chronic liver disease: a prospective study. J Hepatol. 2011;55(5):1017-24.

37.   Pang JXQ, Zimmer S, Niu S, Crotty P, Tracey J, Pradhan F, et al. Liver Stiffness by Transient Elastography Predicts Liver-Related Complications and Mortality in Patients with Chronic Liver Disease. Plos One. 2014;9(4).

38.   Paternostro R, Ferlitsch M, Etschmaier A, Schwarzer R, Reiberger T, Mandorfer M, et al. Transient elastography related cut off parameters in patients with advanced chronic liver disease independently predict decompensation and mortality but lack prediction of varices. Hepatology. 2015;62:925A-6A.

39.   Shen Y, Wu SD, Wu L, Wang SQ, Chen Y, Liu LL, et al. The prognostic role of liver stiffness in patients with chronic liver disease: a systematic review and dose-response meta-analysis. Hepatol Int. 2019;13(5):560-72.

40.   Pons M, Simon-Talero M, Millan L, Ventura-Cots M, Santos B, Augustin S, et al. Basal values and changes of liver stiffness predict the risk of disease progression in compensated advanced chronic liver disease. Dig Liver Dis. 2016;48(10):1214-9.

41.   Kim M, Lim Y, Baik SK, al. e. MELD score and liver stiffness are predictive for the development of acute decompensation that induce acute-on chronic liver failure. . Hepatology. 2015;62:1222A.

42.   Kitson MT, Roberts SK, Colman JC, Paul E, Button P, Kemp W. Liver stiffness and the prediction of clinically significant portal hypertension and portal hypertensive complications. Scandinavian Journal of Gastroenterology. 2015;50(4):462-9.

43.   Vergniol J, Foucher J, Terrebonne E, al. e. Prognostic value of transient elastography and non-invasive markers of fibrosis in patients with chronic liver disease: A prospective analysis of 4,935 person-years. . Hepatology. 2009;50.

44.   Kothari HG, Gupta SJ, Gaikwad NR, Sankalecha TH, Samarth AR. Role of non-invasive markers in prediction of esophageal varices and variceal bleeding in patients of alcoholic liver cirrhosis from central India. The Turkish journal of gastroenterology : the official journal of Turkish Society of Gastroenterology. 2019;30(12):1036-43.

45.   Altman D. Practical statistics for medical research. . London: Chapman and Hall; 1991. 611 p.

46.   Cho EJ, Lee JH, Kim MY, Yoo JJ, Choi WM, Cho YY, et al. Which noninvasive hepatic fibrosis test most effectively predicts portal pressure and survival in alcoholic liver disease patients? Hepatology. 2013;58(4):987A.

47.   Naveau S, Gaude G, Asnacios A, Agostini H, Abella A, Barri-Ova N, et al. Diagnostic and prognostic values of noninvasive biomarkers of fibrosis in patients with alcoholic liver disease. Hepatology. 2009;49(1):97-105.

48.   Raker JM, Sivananthan A, Heald E, Chhaya V, Shalabi A, Forton DM. Defining liver stiffness measurement cut-offs to predict mortality and complications in patients with cirrhosis of mixed etiology. Hepatology. 2016;64 (1 Supplement 1):326A.

49.   Bertrais S, Couffon C, Rondeau L, Favre M, Oberti F, Fouchard-Hubert I, et al. PERFORMANCE OF NON-INVASIVE LIVER FIBROSIS TESTS IN PREDICTING DEATH RATE IN PATIENTS WITH ALCOHOLIC CIRRHOSIS. J Hepatol. 2012;56:S527-S8.

50.   Mueller J, Rausch V, Silva I, Peccerella T, Piecha F, Dietrich C, et al. Survival in a 10 year prospective cohort of heavy drinkers: Liver stiffness is the best long-term prognostic parameter. J Hepatol. 2019;70 (1):e107.

51.   Gomez A, Redondo B, Fraile M, Castano-Garcia A, Torner M, Dieguez MLG, et al. Usefulness of liver stiffness measurement by transient elastography for predicting complications in patients with alcoholic liver disease. J Hepatol.68:S638.

52.   Chaudhari SR, Nigam N, Sharma CB, Rastogi A. Comparison of non-invasive fibrosis estimation scores to predict the outcome in decompensated alcoholic cirrhosis patients. Hepatology International. 2017;11 (1 Supplement 1):S204.

53.   Kothari HG, Gupta SJ, Gaikwad NR, Sankalecha TH, Samarth AR. Role of non-invasive markers in prediction of esophageal varices and variceal bleeding in patients of alcoholic liver cirrhosis from central India. Turkish Journal of Gastroenterology. 2019;30(12):1036-43.

54.   Lee JG, Sohn JH, Jeong JY, Kim TY, Kim SM, Cho YS, et al. Combined Effect of hepatic venous pressure gradient and liver stiffness on long-term mortality in patients with cirrhosis. Journal of Gastroenterology and Hepatology. 2018;33:401-.

55.   Tsochatzis EA, Gurusamy KS, Ntaoula S, Cholongitas E, Davidson BR, Burroughs AK. Elastography for the diagnosis of severity of fibrosis in chronic liver disease: a meta-analysis of diagnostic accuracy. J Hepatol. 2011;54(4):650-9.

56.   Pavlov CS, Casazza G, Nikolova D, Tsochatzis E, Gluud C. Systematic review with meta-analysis: diagnostic accuracy of transient elastography for staging of fibrosis in people with alcoholic liver disease. Aliment Pharmacol Ther. 2016;43(5):575-85.

57.   Crossan C, Tsochatzis EA, Longworth L, Gurusamy K, Papastergiou V, Thalassinos E, et al. Cost-effectiveness of noninvasive liver fibrosis tests for treatment decisions in patients with chronic hepatitis B in the UK: systematic review and economic evaluation. J Viral Hepat. 2016;23(2):139-49.

58.   Poynard T, Ngo Y, Perazzo H, Munteanu M, Lebray P, Moussalli J, et al. Prognostic value of liver fibrosis biomarkers: a meta-analysis. Gastroenterol Hepatol (N Y). 2011;7(7):445-54.

59.   Salkic NN, Jovanovic P, Hauser G, Brcic M. FibroTest/Fibrosure for significant liver fibrosis and cirrhosis in chronic hepatitis B: a meta-analysis. American Journal of Gastroenterology. 2014;109(6):796-809.

60.   Chrostek L, Panasiuk A. Liver fibrosis markers in alcoholic liver disease. World Journal of Gastroenterology. 2014;20(25):8018-23.

61.   Group CM. Cochrane Methods:Prognosis. 'What is prognosis research?' accessed April 2020 [Available from: https://methods.cochrane.org/prognosis/about-us.

62.	Debray TPA, de Jong VMT, Moons KGM, Riley RD. Evidence synthesis in prognosis research. Diagn Progn Res. 2019;3:13.

63.	Steyerberg EW, Moons KG, van der Windt DA, Hayden JA, Perel P, Schroter S, et al. Prognosis Research Strategy (PROGRESS) 3: prognostic model research. PLoS Med. 2013;10(2):e1001381.

64.	Chang Y, Cho YK, Cho J, Jung HS, Yun KE, Ahn J, et al. Alcoholic and Nonalcoholic Fatty Liver Disease and Liver-Related Mortality: A Cohort Study. American Journal of Gastroenterology. 2019;114(4):620-9.

65.	Hyun Kim J, Lee M, Woo Park S, Kang M, Kim M, Hoon Lee S, et al. Validation of modified fibrosis-4 index for predicting hepatocellular carcinoma in patients with compensated alcoholic liver cirrhosis. Medicine (United States). 2018;97(48).

Table 1: Selection criteria

Inclusion criteria	Exclusion Criteria:
All adult humans (age 18+)	Review articles and opinion pieces
Participants have ArLD	Non-human studies
Where studies investigate chronic liver disease of mixed aetiology, they are only to be included if they comprise ≥ 10 patients where the primary aetiology is alcohol, and that these alcohol data are able to be extracted separately.	Pre-clinical and biological studies
Study relates to at least one of the four non-invasive tests of interest (FIB4, ELF, FibroTest, FibroScan)	Aetiology of liver disease other than alcohol
RR, HR or AUROC with corresponding 95% CI must be able to be extracted from the data	Alcoholic hepatitis

RR: Relative Risk, HR: Hazard Ratio, AUROC: Area Under Receiver Operating Characteristic curve, CI: Confidence Interval

Table 2: Baseline characteristics of included studies

Study author, year, location (reference)	Publication type / Retrospective, prospective, or retro-to-prospective	Aetiology	Time period; / Median follow up period (IQR)	Total no pts in study	Total no pts with ArLD included	NIT of interest investigated	Additional prognostic marker assessed	Outcomes assessed (mortality or LRE)	Time point for recorded outcome	Statistical analysis (HR/RR/ AUROC + CI)
Chang et al 2019, South Korea(64)	Full paper / Retro-to-prospective	mixed	2002-2015 (5.2 years, IQR 2.8-8.8)	437,828	17,300	FIB4	APRI	Liver related mortality	End of study period (14 years)	HR + 95%CI
Chaudhari et al 2017, India(52)	Abstract / Not stated	Alcohol only	From Jan 2015 / 7.5 months (5-21)	158	158	FIB4	FIBRO-Q, MELD, APRI, AST:ALT ratio	Mortality (unspecified)	unspecified	AUROC + 95% CI
Raker et al 2016, UK(48)	Abstract / Retrospective	mixed	2008-2014 / 26 months (max 83.6)	408	98	FibroScan	-	All-cause mortality	3 years	AUROC, HR + CI
Bertrais et al 2012, France(49)	Abstract / Prospective	Alcohol only	2004-2009 / 3.4 years (no IQR)	302	302	FibroScan	FibroScan, Fibrometer, Hepascore, CP, Quanti-meter	All-cause mortality and liver-related mortality	1 year	AUROC +95% CI
Mueller et al 2019, Germany(50)	Abstract / Prospective	Alcohol only	2007-2017 / 3.7 years (mean)	943	675	FibroScan	Albumin, bilirubin, ALP, Hb	All-cause mortality	1,3,5 years	AUROC, HR +95% CI

Study		Population	Study period	N	N	Index test	Comparators	Outcome	Follow-up	Statistics
Gomez et al 2018, Spain(51)	Abstract / Prospective	Alcohol only	Not specified / 29.2 years (mean, SD 17.3)	276	276	FibroScan	Child Pugh Score, AST, ALT, platelet count	Liver related event	'outcomes during mean follow-up of 29.2 months (SD 17.3)'	AUROC, OR +95% CI
Hyun Kim et al 2018, South Korea(65)	Full paper / Retrospective	Alcohol only	2007-2015 / 58 months (IQR 31-94)	924	924	FIB4	Modified FIB4, APRI, eLIFT score	Development of HCC	3 years	AUROC +95% CI
Cho E et al 2013, South Korea(46)	Abstract / Not specified	Alcohol only	Not specified / Not specified	195	195	FIB4, FibroScan	APRI, Child Pugh score	Liver-related death and all-cause death	Not specified	AUROC
Naveau et al 2009, France(47)	Full paper / Retro-to-prospective	Alcohol only	Not specified / 8.2 years (range 5 days to 11.8 years)	218 change to 292	218	FibroTest, FIB4	Fibrometer, Hepascore, APRI, Forns'	Liver-related death and all-cause death	Survival at 5 and 10 years	AUROC + 95% CI
Connoley et al, UK, unpublished (15)	Full paper / Prospective	Mixed	Not specified / 6.4 years (IQR 2.8-8.5)	786	64	ELF	Liver biopsy	LRE and all-cause mortality	6 years	HR +95% CI
Kothari et al 2019 (44)	Full paper / Prospective	Alcohol only	2016-2017 / 6 months	202	202	FIB4	APRI, MELD, Child Pugh	Variceal bleed	6 months	AUROC + 95% CI

ALT: Alanine Aminotransferase, AST: Aspartate transaminase, ALP: Alkaline Phosphatase, CP: Child Pugh, OR: Odds Ratio, LRE: Liver-related event, SD: Standard deviation, Hb: Haemoglobin, APRI: AST-to-platelet ratio index, MELD: Model for end-stage liver disease

Table 3: Baseline participant characteristics

Study author, year, location (reference)	Recruitment details (where reported)	Alcohol consumption required for inclusion	Age year mean (SD)	%male	BMI (SD)	ALT (IQR)	% cirrhosis (advanced fibrosis)	Number of events					
								Event rate (incidence rate per 1000 person years)					
								Liver-related mortality	All-cause mortality	HCC	ascites	HE	Variceal bleed
Chang et al 2019(64)	Pt cohort nested in existing multicentre health study. Included pts: those attending employment-related screening clinics with either NAFLD/AFLD based on US and alcohol history. Excluded pts: with evidence of cirrhosis at start of study.	≥30g/day men ≥ 20g/day women	41.6 (9.3)	94.4	26.5 (2.9)	33 (24-48)	0	19	-	-	-	-	-
Chaudhari et al 2017(52)	Inpatients with decompensated alcohol-related cirrhosis	-	43.86 (9.03)	-	-	-	100 % cirrhosis	-	12† — -	-	-	-	-

Study	Population	Criteria											
Raker et al 2016(48)	Inpatients with compensated cirrhosis or advanced fibrosis of mixed aetiology in one UK hospital	-	53.5‡	63‡	-	-	100% either advanced fibrosis or cirrhosis	41 (3yrs) (3%, 6%, 10% at 1, 2, and 3 yrs)‡ —	-	-	-	-
								-					
Bertrais et al 2012(49)	Patients with alcohol-related cirrhosis with no history of HCC	-	60	69.9	-	-	100% cirrhosis	50 —	91 —	-	-	-	-
								-	-				
Mueller et al 2019,(50)	Caucasian heavy drinkers that had presented for alcohol detoxification (6 days)	presented for alcohol detoxification) (Mean consumption was 178 g/d)	-	-	-	-	-	16 —	106 —	-	-	-	-
								-	-				
Gomez et al 2018, Spain (51)	Patients with Child Pugh A/B alcohol-related cirrhosis without HCC or decompensation at time of enrolment	-	56.5 (8.4)	82	-	-	100% (of which 80% child A, 20% child B)	-	-	13	29	14	17
Hyun et al 2018(65)	Inpatients and outpatients with alcohol-related cirrhosis, excluding 'active alcoholism' and excluding decompensation or HCC at enrolment	Alcohol consumption >10yrs, exceeding 60g/day for males, 40g/day females, but no alcohol for past 2 yrs	59	62.9	-	19	100%	-	-	- —	-	-	-
										-			

Cho E et al 2013(46)	'patients with alcohol related liver disease'	-		-	-	-	-	-	-	-	-	-	-	-
Naveau et al 2009(47)	'patients with heavy alcohol consumption and available liver biopsy and FibroTest results'	Patients had to have consumed at least 50g of alcohol per day over past year	47 (0.7)	78	-	65 (SD 5)	31% cirrhosis (biopsy)	42 -	85 -	7 -	-	-	4 -
Connoley et al (unpublished)(15)	Patients aged between 18-75 yrs undergoing a planned liver biopsy	-	50 (IQR 41.5-57.5)	67.9	-	36 (23-66)	77.8% ≥F3, 66.7% ≥F5 (biopsy)	32 at 6 yrs, 34 at 7 yrs, 35 at 8 yrs	23 at 6 yrs, 26 at 7 yrs, 26 at 8 years	-	-	-	-
Kothari et al 2019 (44)	Male patients aged 18-70 with clinical diagnosis of Alcohol-related cirrhosis, absence of TIPS/previous variceal bleed	'clinically significant alcohol intake'	43.77 (9.95)	100	-	29 (21-50)	100% (clinical/ imaging-based diagnosis)	-	-	-	-	-	61

† Presumed all-cause mortality not liver-related, but not actually specified in study

‡ Of whole study cohort. (Data on this not reported for alcohol cohort separately)

HE: Hepatic Encephalopathy, HCC: Hepatocellular carcinoma, RR: relative risk, HR: Hazard ratio, BMI: body mass index, ALT: alanine aminotransferase, NAFLD: non-alcoholic fatty liver disease, AFLD: alcoholic fatty liver diseae

Table 4: Prognostic performance of NIT in each study

FIB4										
Study	**Cut off or continuous NIT value**	**outcome**	**AUROC**	**95% CI**		**HR**	**95% CI**		**Other analysis**	**Adjustment for confounding factors**
				Lower limit	**Upper limit**		**Lower limit**	**Upper limit**		
Chang et al (64)	Low FIB 4 = <1.3 Intermediate FIB4 = 1.3 to <2.67 High FIB4 = ≥ 2.67	LRM (14 years)	- - -	- - -	- - -	1.14 4.48 32.9	0.34 1.91 15.04	3.85 10.5 71.96		Yes (sex, yr of screening exam, center, BMI, smoking status, regular exercise, educational level, diabetes, HTN)
Chaudhary et al (52)	Not stated	Mortality (unspecified if LRM, unspecified time point)	0.825	0.71	0.93	-	-	-		Not stated
Hyun et al 2018 (65)	Continuous FIB4 and cut offs – low FIB4 ≤3.25, high FIB4 >3.25	HCC (3 years)	0.69	0.63	0.75	-	-	-	Fib4 high versus low HR 1.71 (95% CI 1.08-2.71)	Yes (age, albumin, platelets, modified FIB4, APRI, eLIFT score)
Cho E et al 2013 (46)	Continuous FIB4	LRM (unspecified time point)	0.78	-	-	1.11	-	-	-	Yes (age)
Naveau et al (47)	Continuous FIB4	ACM (unspecified time point)	0.64	0.55	0.71	-	-	-	-	No

Study	Test	Outcome	AUC	CI low	CI high	HR	HR low	HR high	Threshold comparison	Adjusted
Kothari et al 2019 (44)	Continuous FIB4	Index variceal bleed (6 months)	0.74	0.66	0.81	-	-	-	-	No
FibroScan										
Cho E et al 2013 (46)	Continuous liver stiffness (kPa)	LRM (unspecified time point)	0.73	-	-	-	-	-	-	Yes (Age)
Raker et al (48)	Continuous LSM + Threshold of <20kpa vs >20kpa	ACM (3 yrs)	0.74	-	-	-†	-	-	Highest vs lowest – 3% incidence of death with LSM <20kpa, VS 15% deaths with LSM >20kpa (p = <0.004)	No
Bertrais et al (49)	Continuous LSM	ACM (1-yr) LRM (1yr)	0.65 0.73	0.51 0.64	0.79 0.83	-	-	-	-	no
Mueller et al (50)	Continuous LSM	ACM (1-yr) ACM (3-yr) ACM (5-yr)	0.76 0.74 0.73	-	-	1.013	1.003	1.023	-	Yes (Age, ALP, albumin)
Gomez et al (51)	Continuous LSM +threshold of <25kpa vs >25kpa	LRE (during follow up period) (unspecified time point)	0.675	0.607	0.743	-	-	-	Highest vs lowest (TE <25kPa = mean incidence of 4.5% for LRE versus 15.5% for >25kPa	Yes (sex, Child Pugh score)
FibroTest										
Naveau et al (47)	Continuous FibroTest value Continuous FibroTest value	LRM (unspecified time point)	0.79 0.69	0.68 0.61	0.86 0.76	23.2‡ 3.7‡	3.2 1.2	167.3 11.7		Yes (liver biopsy, fibrometer, hepascore,

		ACM (unspecified time point)								abstinent vs non-abstinent
	FibroTest cut offs:								5-yr SNLRD	
	0-0.31 (no or minimal fibrosis)	SNLRD (5-yr)							98.7% (96-1)	
	0.21-0.58 (moderate fibrosis)	SNLRD (5-yr)							92.1% (83.5-100)	
	0.59-1 (severe fibrosis)	SNLRD (5-yr)							68.3% (79.5-89.4)	
	FibroTest cut offs:								10-yr SNLRD	
	0-0.31 (no or minimal fibrosis)	SNLRD (10-yr)							92% (84.9-99)	
	0.21-0.58 (moderate fibrosis)	SNLRD (10-yr)							87.5% (75.5-99.5)	
	0.59-1 (severe fibrosis)	SNLRD (10-yr)							78.5% (72.4-84.6)	
ELF										
Connoley et al (unpublished) (15)	ELF as continuous	LRE (6-yr)	-	-	-	1.82§	1.169	2.83	-	Yes (age and sex)
	ELF as continuous	LRE (6-yr)	0.816	0.713	0.920	-	-	-	-	no
	ELF as continuous	LRE (7-yr)	0.844	0.750	0.938					no
	ELF as continuous	LRE (8-yr)	0.847	0.754	0.940					no
	ELF as continuous	ACM (6-yr)	0.733	0.645	0.861	-	-	-	-	no
	ELF as continuous	ACM (7-yr)	0.722	0.591	0.852	-	-	-	-	no
	ELF as continuous	ACM (8-yr)	0.722	0.591	0.852	-	-	-	-	no

ELF cut offs in 4 categories (compared to <9.8)		-	-	-				-	Yes (age and sex)
9.8-10.49	LRE (6-yr)				1.49	0.287	7.74		
10.5-11.29	LRE (6-yr)				3.84	0.9	16.39		
≥11.3	LRE (6-yr)				10.24	2.97	35.27		
ELF cut offs in two categories		-	-	-				-	Yes (age and sex)
<10.5 and ≥10.5	LRE (6-yr)				6.42	2.63	15.24		

† Study reported HR with corresponding 95% CI for LSM cut offs above and below 20kPa, and for ArLD versus non-ArLD, but not specifically for LSM in the ArLD cohort.

‡ RR (risk ratio) not HR

§ OR (odds ratio)

LRM = Liver Related Mortality, LRE: Liver related event, ACM: all-cause mortality, SNLRD: Survival or Non-liver related Death, LSM: Liver Stiffness Measurement, HTN: hypertension, TE: transient elastography

Identification

Records identified through database searching
(n =25,088)
(of which, WOS= 9,653, Medline and EMBASE =
15,424, Cochrane = 11 (3 Cochrane reviews, 8
Cochrane trials)
8781 duplicates)

Additional records identified through
other sources
(reference list search from included
papers, n = 8)
(contacting author for unpublished
paper n = 1)

Records after duplicates removed
(n = 16,307 + 9 from other sources)

Screening

Records screened by 1ˢᵗ
reviewer by title/abstract
(n = 16,316)

Records excluded
(n = 15,296)

Abstracts screened by 2 reviewers
(n = 1020) (1011 +8+1)

Records excluded
(n = 982)

Eligibility

Full-text articles assessed
by 2 reviewers for
eligibility (n = 40)

Full-text articles excluded, with
reasons (n = 29)
1. Data on CLD, with alcohol
   data not reported
   separately (n=14)
2. Wrong outcome (n=1)
3. Aetiology of CLD
   undefined or unclear if
   includes alcohol (n= 8)
4. Inappropriate/insufficient
   outcome reporting (n=2)
5. Insufficient number of
   patients with ArLD (n=<10)
   (n= 1)
6. Wrong study design (n=1)
7. Pt cohort represented in
   another more recent and
   comprehensive included
   study (n=2)

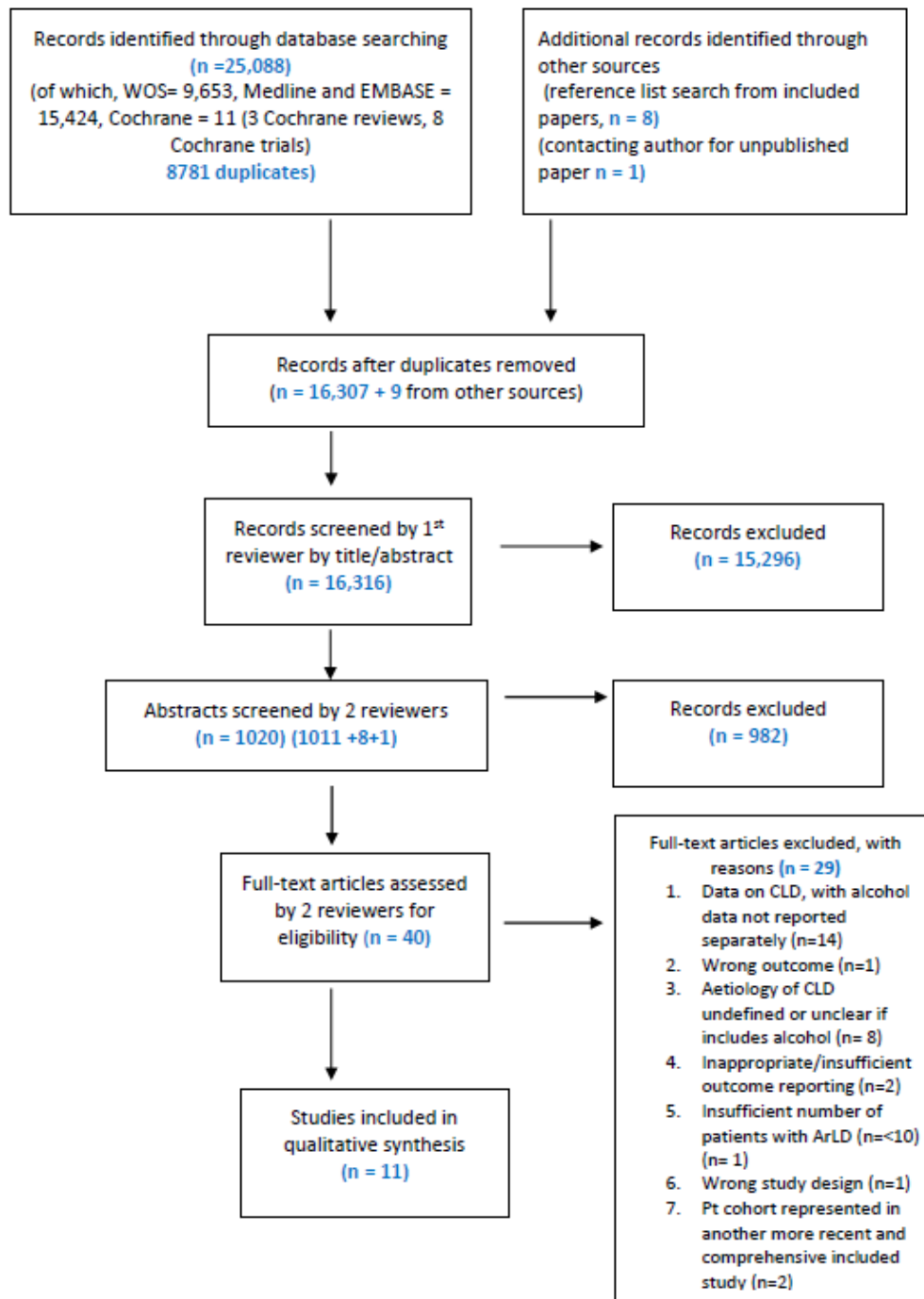**Included**

Studies included in
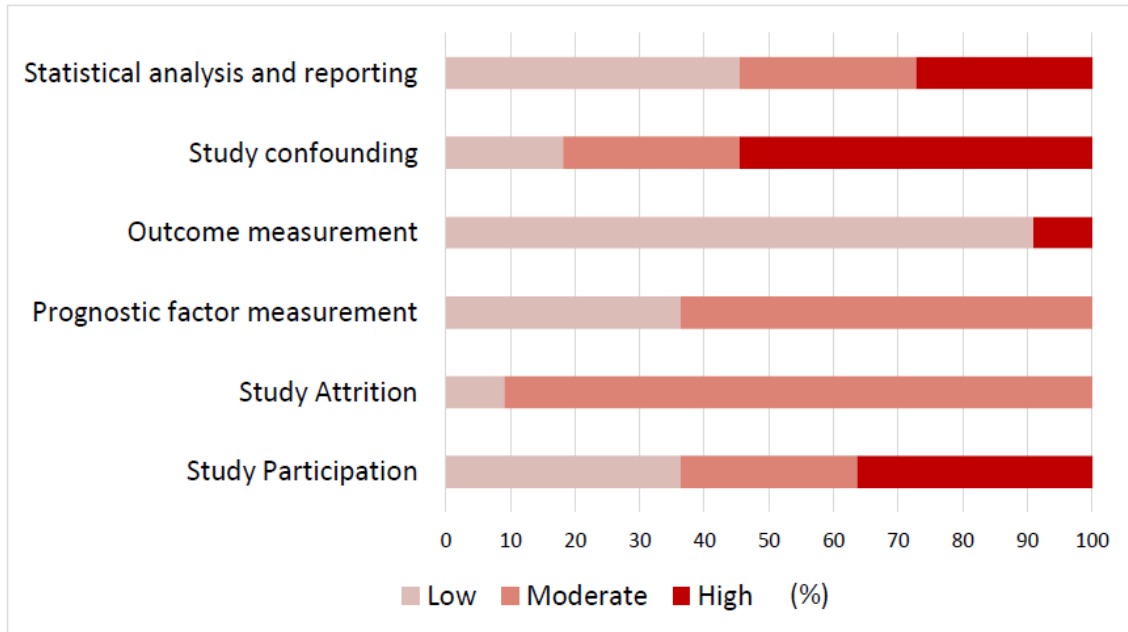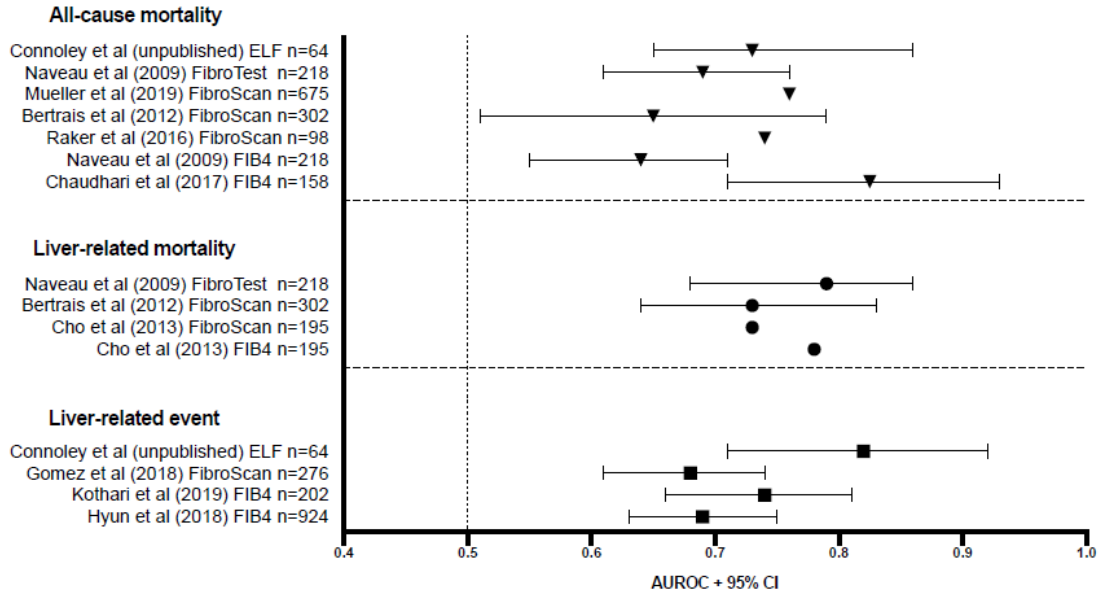qualitative synthesis
(n = 11)

FIGURE 1: PRISMA FLOW DIAGRAM.

Figure 2: QUIPs tool results: Quality assessment of included studies using the quality in prognosis studies tool.

All-cause mortality
Connoley et al (unpublished) ELF n=64
Naveau et al (2009) FibroTest n=218
Mueller et al (2019) FibroScan n=675
Bertrais et al (2012) FibroScan n=302
Raker et al (2016) FibroScan n=98
Naveau et al (2009) FIB4 n=218
Chaudhari et al (2017) FIB4 n=158

Liver-related mortality
Naveau et al (2009) FibroTest n=218
Bertrais et al (2012) FibroScan n=302
Cho et al (2013) FibroScan n=195
Cho et al (2013) FIB4 n=195

Liver-related event
Connoley et al (unpublished) ELF n=64
Gomez et al (2018) FibroScan n=276
Kothari et al (2019) FIB4 n=202
Hyun et al (2018) FIB4 n=924

AUROC + 95% CI

* Chang et al 2019 study on FIB4 not represented in this forest plot as study did not report AUROCs, however, this study had largest sample size of 17,300. HR for liver-related mortality at 14 years = Low FIB4 threshold <1.3: HR 1.14 (95% CI 0.34-3.85), Intermediate FIB4 1.3-2.67: HR 4.48 (95% CI 1.91-10.5), High FIB4 ≥ 2.67: HR 32.9 (95% CI 15.04-71.96).

Figure 3: Forest plot of AUROCs + 95% CI for outcome prediction*