Cochrane Library

Cochrane Database of Systematic Reviews

Liver fibrosis stage based on the four factors (FIB-4) score or Forns index in adults with chronic hepatitis C (Review)

Huttman M, Parigi TL, Zoncapè M, Liguori A, Kalafateli M, Noel-Storr AH, Casazza G, Tsochatzis E

Huttman M, Parigi TL, Zoncapè M, Liguori A, Kalafateli M, Noel-Storr AH, Casazza G, Tsochatzis E.
Liver fibrosis stage based on the four factors (FIB-4) score or Forns index in adults with chronic hepatitis C.
Cochrane Database of Systematic Reviews 2024, Issue 8. Art. No.: CD011929.
DOI: 10.1002/14651858.CD011929.pub2.

www.cochranelibrary.com

WILEY

TABLE OF CONTENTS

[Diagnostic Test Accuracy Review]

Liver fibrosis stage based on the four factors (FIB-4) score or Forns index in adults with chronic hepatitis C

Marc Huttman[1], Tommaso Lorenzo Parigi[1], Mirko Zoncapè[1], Antonio Liguori[1], Maria Kalafateli[1], Anna H Noel-Storr[2], Giovanni Casazza[3,4], Emmanuel Tsochatzis[1]

[1]Sheila Sherlock Liver Centre, Royal Free Hospital and the UCL Institute of Liver and Digestive Health, London, UK. [2]Radcliffe Department of Medicine, University of Oxford, Oxford, UK. [3]Department of Clinical Sciences and Community Health - Laboratory of Medical Statistics, Biometry and Epidemiology "G.A. Maccacaro", Università degli Studi di Milano, Milan, Italy. [4]Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, Milan, Italy

Contact: Emmanuel Tsochatzis, e.tsochatzis@ucl.ac.uk.

Editorial group: Cochrane Hepato-Biliary Group.
Publication status and date: New, published in Issue 8, 2024.

Citation: Huttman M, Parigi TL, Zoncapè M, Liguori A, Kalafateli M, Noel-Storr AH, Casazza G, Tsochatzis E. Liver fibrosis stage based on the four factors (FIB-4) score or Forns index in adults with chronic hepatitis C. *Cochrane Database of Systematic Reviews* 2024, Issue 8. Art. No.: CD011929. DOI: 10.1002/14651858.CD011929.pub2.

ABSTRACT

Background

The presence and severity of liver fibrosis are important prognostic variables when evaluating people with chronic hepatitis C (CHC). Although liver biopsy remains the reference standard, non-invasive serological markers, such as the four factors (FIB-4) score and the Forns index, can also be used to stage liver fibrosis.

Objectives

To determine the diagnostic accuracy of the FIB-4 score and Forns index in staging liver fibrosis in people with chronic hepatitis C (CHC) virus, using liver biopsy as the reference standard (primary objective). To compare the diagnostic accuracy of these tests for staging liver fibrosis in people with CHC and explore potential sources of heterogeneity (secondary objectives).

Search methods

We used standard Cochrane search methods for diagnostic accuracy studies (search date: 13 April 2022).

Selection criteria

We included diagnostic cross-sectional or case-control studies that evaluated the performance of the FIB-4 score, the Forns index, or both, against liver biopsy, in the assessment of liver fibrosis in participants with CHC. We imposed no language restrictions. We excluded studies in which: participants had causes of liver disease besides CHC; participants had successfully been treated for CHC; or the interval between the index test and liver biopsy exceeded six months.

Data collection and analysis

Two review authors independently extracted data. We performed meta-analyses using the bivariate model and calculated summary estimates. We evaluated the performance of both tests for three target conditions: significant fibrosis or worse (METAVIR stage ≥ F2); severe fibrosis or worse (METAVIR stage ≥ F3); and cirrhosis (METAVIR stage F4). We restricted the meta-analysis to studies reporting cut-offs in a specified range (+/-0.15 for FIB-4; +/-0.3 for Forns index) around the original validated cut-offs (1.45 and 3.25 for FIB-4; 4.2 and 6.9 for Forns index). We calculated the percentage of people who would receive an indeterminate result (i.e. above the rule-out threshold but below the rule-in threshold) for each index test/cut-off/target condition combination.

Main results

We included 84 studies (with a total of 107,583 participants) from 28 countries, published between 2002 and 2021, in the qualitative synthesis. Of the 84 studies, 82 (98%) were cross-sectional diagnostic accuracy studies with cohort-based sampling, and the remaining two (2%) were case-control studies. All studies were conducted in referral centres. Our main meta-analysis included 62 studies (100,605 participants).

Overall, two studies (2%) had low risk of bias, 23 studies (27%) had unclear risk of bias, and 59 studies (73%) had high risk of bias. We judged 13 studies (15%) to have applicability concerns regarding participant selection.

FIB-4 score

The FIB-4 score's low cut-off (1.45) is designed to rule out people with at least severe fibrosis (≥ F3). Thirty-nine study cohorts (86,907 participants) yielded a summary sensitivity of 81.1% (95% confidence interval (CI) 75.6% to 85.6%), specificity of 62.3% (95% CI 57.4% to 66.9%), and negative likelihood ratio (LR-) of 0.30 (95% CI 0.24 to 0.38).

The FIB-4 score's high cut-off (3.25) is designed to rule in people with at least severe fibrosis (≥ F3). Twenty-four study cohorts (81,350 participants) yielded a summary sensitivity of 41.4% (95% CI 33.0% to 50.4%), specificity of 92.6% (95% CI 89.5% to 94.9%), and positive likelihood ratio (LR+) of 5.6 (95% CI 4.4 to 7.1).

Using the FIB-4 score to assess severe fibrosis and applying both cut-offs together, 30.9% of people would obtain an indeterminate result, requiring further investigations. We report the summary accuracy estimates for the FIB-4 score when used for assessing significant fibrosis (≥ F2) and cirrhosis (F4) in the main review text.

Forns index

The Forns index's low cut-off (4.2) is designed to rule out people with at least significant fibrosis (≥ F2). Seventeen study cohorts (4354 participants) yielded a summary sensitivity of 84.7% (95% CI 77.9% to 89.7%), specificity of 47.9% (95% CI 38.6% to 57.3%), and LR- of 0.32 (95% CI 0.25 to 0.41).

The Forns index's high cut-off (6.9) is designed to rule in people with at least significant fibrosis (≥ F2). Twelve study cohorts (3245 participants) yielded a summary sensitivity of 34.1% (95% CI 26.4% to 42.8%), specificity of 97.3% (95% CI 92.9% to 99.0%), and LR+ of 12.5 (95% CI 5.7 to 27.2).

Using the Forns index to assess significant fibrosis and applying both cut-offs together, 44.8% of people would obtain an indeterminate result, requiring further investigations. We report the summary accuracy estimates for the Forns index when used for assessing severe fibrosis (≥ F3) and cirrhosis (F4) in the main text.

Comparing FIB-4 to Forns index

There were insufficient studies to meta-analyse the performance of the Forns index for diagnosing severe fibrosis and cirrhosis. Therefore, comparisons of the two tests' performance were not possible for these target conditions. For diagnosing significant fibrosis and worse, there were no significant differences in their performance when using the high cut-off. The Forns index performed slightly better than FIB-4 when using the low/rule-out cut-off (relative sensitivity 1.12, 95% CI 1.00 to 1.25; P = 0.0573; relative specificity 0.69, 95% CI 0.57 to 0.84; P = 0.002).

Authors' conclusions

Both the FIB-4 score and the Forns index may be considered for the initial assessment of people with CHC. The FIB-4 score's low cut-off (1.45) can be used to rule out people with at least severe fibrosis (≥ F3) and cirrhosis (F4). The Forns index's high cut-off (6.9) can be used to diagnose people with at least significant fibrosis (≥ F2). We judged most of the included studies to be at unclear or high risk of bias. The overall quality of the body of evidence was low or very low, and more high-quality studies are needed. Our review only captured data from referral centres. Therefore, when generalising our results to a primary care population, the probability of false positives will likely be higher and false negatives will likely be lower. More research is needed in sub-Saharan Africa, since these tests may be of value in such resource-poor settings.

PLAIN LANGUAGE SUMMARY

How accurate are the FIB-4 score and Forns index (non-invasive tests) in diagnosing liver fibrosis (scarring) stages in adults with chronic hepatitis C?

Key messages

• Both the FIB-4 score and Forns index can be used in the initial phase of investigating whether someone has liver scarring.

• It is best to use the FIB-4 score to rule out stage 3 (severe fibrosis) or stage 4 scarring (cirrhosis).

• It is best to use the Forns index to diagnose people with stage 2 scarring (significant fibrosis).

Why is improving the diagnosis of liver scarring important?

Hepatitis C infection is a common cause of liver scarring (fibrosis). Untreated, liver scarring can progress to a severe form called liver cirrhosis, which is mostly irreversible and can cause the liver to shut down or develop cancer. Currently, the best test to diagnose liver fibrosis is liver biopsy, where liver tissue is taken with a needle and looked at under a microscope. However, liver biopsy is invasive, costly, painful, and carries some serious risks such as bleeding. Accurately diagnosing liver fibrosis through non-invasive tests such as the FIB-4 score and Forns index would benefit people and healthcare systems overall. However, their diagnostic accuracy (that is, how good they are at telling us which people have what stage of disease) in people with hepatitis C infection remains unclear.

What are the FIB-4 score and Forns index tests?

The FIB-4 score and Forns index are tests for diagnosing stages of liver fibrosis. They combine standard laboratory results with factors such as age to calculate a score that estimates the amount of scarring in the liver. Compared to liver biopsy, these are simple, inexpensive, widely available, relatively painless, and risk-free tests.

Each test has two cut-offs: high/rule in and low/rule out. If a person's result is below the low cut-off, they *do not* have that stage of fibrosis. If a person's result is above the high cut-off, they *do* have that stage of fibrosis. If someone's score is between the two cut-offs, the test is unhelpful because it can neither rule in nor rule out fibrosis. This is called the 'grey area'. Someone with a score in the 'grey area' should have further tests, such as a liver biopsy.

What did we want to find out?

We wanted to determine how well the FIB-4 score and Forns index can diagnose different liver fibrosis stages in people with chronic hepatitis C, compared to the results from liver biopsy.

What did we do?

We searched for studies that evaluated the diagnostic accuracy of the FIB-4 score or Forns index (or both) in people with hepatitis C. We combined the results from these studies.

What did we find?

We included 84 studies with a total of 107,583 participants. The studies were conducted in 28 countries, and were published between 2002 and 2021. We analysed results from 62 studies with 100,605 participants. We selected this portion of studies because they applied the two tests using comparable low and high cut-off values. This approach means we can be more confident about the results of our analysis.

By combining the studies' results for the FIB-4 score for diagnosing severe (stage 3) fibrosis, we can say the following for a hypothetical group of 1000 people:

• using the high or 'rule-in' cut-off, 144 people would correctly be diagnosed with severe fibrosis, whilst 48 people would wrongly be diagnosed with this stage of disease;
• using the low or 'rule-out' cut-off, 430 people would have severe fibrosis correctly ruled out, whilst 58 people *with* fibrosis would be missed;
• by using both cut-offs together, about one-third of people will need further tests ('grey area').

By combining the studies' results for the Forns index to diagnose significant (stage 2) fibrosis, we can say the following for a hypothetical group of 1000 people:

• using the high or 'rule-in' cut-off, 179 people would correctly be diagnosed with significant fibrosis, whilst 13 people would wrongly be diagnosed with this stage of disease;
• using the low or 'rule-out' cut-off, 218 people would have significant fibrosis correctly ruled out, whilst 83 people would be missed;
• by using both cut-offs together, about half of people will need further tests ('grey area').

What are the limitations of the evidence?

Our confidence in the evidence was reduced because many of the studies may have overestimated the diagnostic accuracy of the tests. Also, the numbers described above are a summary based on pooling results from many studies. Because estimates of accuracy varied considerably across individual studies, we cannot be sure that applying the FIB-4 score or Forns index will always produce these results.

How up to date is this evidence?

The evidence is current to 13 April 2022.

SUMMARY OF FINDINGS

Summary of findings 1. Diagnostic accuracy of the FIB-4 score and Forns index for diagnosing liver fibrosis in people with chronic hepatitis C infection

Review details
Review question
What is the diagnostic accuracy of the FIB-4 score and the Forns index for the diagnosis of different stages of liver fibrosis in people with chronic hepatitis C infection?
Population
Adults diagnosed with chronic hepatitis C, with or without HIV co-infection. Other aetiologies of liver disease such as alcohol abuse, hepatitis B, and schistosomiasis were excluded.
Setting
Any clinical setting (primary, secondary or tertiary)
Index tests
• The FIB-4 score is a non-invasive test combining standard biochemical values with age. The formula is: age ([yr] x AST [U/L]) / ((PLT [10^9/L]) x (ALT$^{1/2}$ [U/L])). • The Forns index is another non-invasive test using different standard biochemical values. The formula is: 7.811 − 3.131 × ln(PLT) + 0.781 × ln(GGT) + 3.467 × ln(age) − 0.014 × (total cholesterol).
Target conditions
The target conditions were the presence of three different stages of hepatic fibrosis in the selected population, as defined by the METAVIR stage: • participants with at least significant fibrosis (stages F2 to F4) compared to participants with no significant fibrosis (stages F0 to F1); • participants with at least severe fibrosis (stages F3 to F4) compared to participants without severe fibrosis (stages F0 to F2); • participants with cirrhosis (stage F4) compared to participants with no cirrhosis (stages F0 to F3).
Role and purpose of index tests
The FIB-4 score and the Forns index are both indirect non-invasive markers of liver fibrosis based on laboratory values, and have been developed as an alternative to liver biopsy in detecting and staging liver fibrosis and diagnosing cirrhosis.
Reference standard
The reference standard used was histopathological examination of liver tissue (liver biopsy) obtained through percutaneous, transjugular, or laparoscopic biopsy.
Limitations in the evidence
Risk of bias in 74 studies reporting on the diagnostic accuracy of the FIB-4 score

- Participant selection: 4 studies (5%) high risk, 26 studies (35%) unclear risk, 44 studies (59%) low risk
- Index test (FIB-4): 39 studies (53%) high risk, 35 studies (47%) low risk
- Reference standard: 9 studies (12%) high risk, 57 studies (77%) unclear risk, 8 studies (11%) low risk
- Flow and timing: 16 studies (22%) high risk, 23 studies (31%) unclear risk, 35 studies (47%) low risk
- Overall: 53 studies (72%) high risk, 21 studies (28%) unclear risk, 0 studies (0%) low risk

Risk of bias in 25 studies reporting on the diagnostic accuracy of the Forns index

- Participant selection: 1 study (4%) high risk, 2 studies (8%) unclear risk, 22 studies (88%) low risk
- Index test (Forns index): 9 studies (36%) high risk, 16 studies (64%) low risk
- Reference standard: 5 studies (20%) high risk, 15 studies (60%) unclear risk, 5 studies (20%) low risk
- Flow and timing: 6 studies (24%) high risk, 5 studies (20%) unclear risk, 14 studies (56%) low risk
- Overall: 16 studies (64%) high risk, 7 studies (28%) unclear risk, 2 studies (8%) low risk

Applicability concerns

- In the 74 studies reporting on the FIB-4 score: 11 studies (15%) had applicability concerns relating to participant selection, 63 studies (85%) had no applicability concerns.
- In the 25 studies reporting on the Forns index: 2 studies (8%) had applicability concerns relating to participant selection, 23 studies (92%) had no applicability concerns.

Key findings (implications in a hypothetical cohort of 1000 people with chronic hepatitis C infection)

This table examines the diagnostic accuracy of the FIB-4 score and Forns index for the target conditions they were originally designed for (i.e. FIB-4 to rule in/out severe fibrosis ≥ F3; Forns index to rule in/out significant fibrosis ≥ F2).

Where the prevalence of the target condition was required to calculate items within this table, we used the median prevalence in the included studies.

Index test	Target condition	Cut-off used	Number of study cohorts (participants)	Sensitivity (95% CI)	Specificity (95% CI)	Median prevalence (IQR) %	Likelihood ratios (95% CI)	True positives	False negatives	True negatives	False positives	Applying both cut-offs together	Certainty of the evidence
FIB-4	Severe fibrosis (≥ F3)	1.45 - low/ rule out cut-off	39 (86,907)	81.1% (75.6% to 85.6%)	62.3% (57.4% to 66.9%)	30.9 (23.8 to 43.9)	LR- 0.30 (0.24 to 0.38)	251 people with fibrosis test positive and are appropriately prioritised for antiviral treatment.	58 people with fibrosis test negative and are not prioritised for antiviral treatment.	430 people without fibrosis test negative and are appropriately not prioritised for antiviral treatment.	261 people without fibrosis test positive and are inappropriately prioritised for antiviral treatment.	By applying the low/rule out cut-off, and the high/rule in cut-off together, 309 people would obtain a 'grey-area' result between 1.45 and 3.25,	Very low[a] ⓧ◯◯◯

		3.25 - high/ rule in cut-off	24 (81,350)	41.4% (33.0% to 50.4%)	92.6% (89.5% to 94.9%)	34.8 (19.1 to 42.5)	LR+ 5.6 (4.4 to 7.1)	144 people with fibrosis test positive and are appropriately prioritised for antiviral treatment.	204 people with fibrosis test negative and are not prioritised for antiviral treatment.	604 people without fibrosis test negative and are appropriately not prioritised for antiviral treatment.	48 people without fibrosis test positive and are inappropriately prioritised for antiviral treatment.	and require further testing. (N.B. calculated from the subset of 18 studies that reported both cut-offs together)	Low[b] ⓍⓍ○○
Forns index	Significant fibrosis (≥ F2)	4.2 - low/ rule out cut-off	17 (4354)	84.7% (77.9% to 89.7%)	47.9% (38.6% to 57.3%)	54.5 (40.0 to 62.0)	LR- 0.32 (0.25 to 0.41)	462 people with fibrosis test positive and are appropriately prioritised for antiviral treatment.	83 people with fibrosis test negative and are not prioritised for antiviral treatment.	218 people without fibrosis test negative and are appropriately not prioritised for antiviral treatment.	237 people without fibrosis test positive and are inappropriately prioritised for antiviral treatment.	By applying the low/rule out cut-off, and the high/rule in cut-off together, 448 people would obtain a 'grey area' result between 4.2 and 6.9, and require further testing. (N.B. calculated from the subset of 12 studies that reported both cut-offs together)	Low[b] ⓍⓍ○○
		6.9 - high/ rule in cut-off	12 (3245)	34.1% (26.4% to 42.8%)	97.3% (92.9% to 99.0%)	52.5 (34.2 to 58.2)	LR+ 12.5 (5.7 to 27.2)	179 people with fibrosis test positive and are appropriately prioritised for antiviral treatment.	346 people with fibrosis test negative and are not prioritised for antiviral treatment.	462 people without fibrosis test negative and are appropriately not prioritised for antiviral treatment.	13 people without fibrosis test positive and are inappropriately prioritised for antiviral treatment.		Very low[c] Ⓧ○○○

CAUTION: the results in this table should be interpreted in conjunction with the results of the individual studies contributing to each summary test accuracy measure. We report these results in the main body of the review. The proportions of true positives, true negatives, false positives, and false negatives in specific settings will be affected by local disease prevalence.

ALT: alanine transaminase; **AST**: aspartate aminotransferase; **CI:** confidence interval; **GGT**: gamma-glutamyl transferase; **IQR:** interquartile range; **ln:** natural logarithm; **LR +:** positive likelihood ratio; **LR−:** negative likelihood ratio; **PLT**: platelet count

[a]Downgraded by one level for risk of bias due to high number of studies with an unclear risk of bias. No downgrades for imprecision as 95% CIs for summary estimates are narrow enough that clinical action would not differ if the upper versus the lower boundary of the CI represented the truth. Downgraded by two levels for inconsistency due to extreme dissimilarity between point estimates of individual studies. No downgrades for indirectness as there were low applicability concerns.

[b]Downgraded by one level for risk of bias due to high number of studies with an unclear risk of bias. No downgrades for imprecision as 95% CIs for summary estimates are narrow enough that clinical action would not differ if the upper versus the lower boundary of the CI represented the truth. Downgraded by one level for inconsistency due to dissimilarity between point estimates of individual studies. No downgrades for indirectness as there were low applicability concerns.

[c]Downgraded by one level for risk of bias due to high number of studies with an unclear risk of bias. Downgraded by one level for imprecision as 95% CIs for summary estimates are wide enough that clinical action would differ if the upper versus the lower boundary of the CI represented the truth (i.e. a true LR+ 27.2 would confer very different clinical significance to LR+ 5.7). Downgraded by one level for inconsistency due to dissimilarity between point estimates of individual studies. No downgrades for indirectness as there were low applicability concerns.

BACKGROUND

Chronic hepatitis C is a major cause of chronic liver disease, with approximately 71 million individuals infected worldwide. On average, 10% to 20% of infected people develop cirrhosis (widespread scarring and impaired function of the liver), over 20 to 30 years of infection. Subsequently, 1% to 5% of people with hepatitis C virus-related cirrhosis will develop hepatocellular carcinoma (liver cancer). Once a diagnosis of hepatocellular carcinoma is established, the probability of death during the first year is 33% (EASL 2015). Chronic hepatitis C is diagnosed by finding both hepatitis C virus antibodies and hepatitis C virus ribonucleic acid (RNA) in the blood, along with signs of ongoing liver inflammation, such as elevated aminotransferase levels or liver tissue changes seen on biopsy.

The extent of liver damage caused by chronic hepatitis C varies from minimal changes to significant fibrosis (scarring) and cirrhosis. According to the 2020 European Association for the Study of the Liver (EASL) Clinical Practice Guidelines, liver disease severity in chronic hepatitis C needs to be assessed before the initiation of antiviral treatment (EASL 2020). Following this, the mainstay of modern treatment for people with a confirmed chronic hepatitis C viral (HCV) infection is direct-acting antiviral agents (DAA). The ultimate goal of this treatment is to achieve undetectable HCV-RNA viral loads 12 or 24 weeks after the termination of treatment, termed ‘sustained virological response’ (SVR) (EASL 2020).

Current guidelines state that everyone with HCV should receive antiviral treatment, irrespective of their stage of liver fibrosis (EASL 2021). However, diagnosing someone's stage of fibrosis in HCV infection is important as it guides multiple facets of management, including treatment choice, additional screening, treatment setting, and follow-up. Thus, for example, people without cirrhosis or with compensated cirrhosis (where the liver can still perform most essential functions, despite extensive scarring and other changes) should be treated with DAA regimens that are pangenotypic (i.e. effective against all strains of hepatitis C virus) and free from interferon and ribavirin (drugs traditionally used to treat HCV that have significant adverse effects). Such regimens include giving pangenotypic, interferon, and ribavirin-free regimens such as sofosbuvir and velpatasvir for 12 weeks, or glecaprevir and pibrentasvir for eight weeks (EASL 2020). People with decompensated cirrhosis should also start on ribavirin daily, or if not tolerated, a longer 24-week course of sofosbuvir and velpatasvir (EASL 2020). Staging liver disease also allows planning of additional screening and follow-up. For example, people diagnosed with HCV-related cirrhosis should be screened for hepatocellular carcinoma and undergo endoscopic screening for oesophageal or gastric varices (abnormally enlarged veins in the oesophagus or stomach that may rupture and bleed) upon diagnosis (Tsochatzis 2014). Regarding follow-up, people with advanced fibrosis or cirrhosis who have achieved sustained virological response must be offered dedicated follow-up after treatment, for screening of hepatocellular carcinoma and provision of further care (EASL 2020). In contrast, people with no to moderate fibrosis with sustained virological response can be discharged, provided they have no comorbidities (EASL 2020).

Liver biopsy methods

A liver biopsy is a procedure to obtain a small sample of liver tissue for diagnostic purposes. Since the mid-1960s, liver biopsy has been considered the 'best reference standard' for the diagnosis and staging of liver fibrosis (Standish 2006).
The histopathological examination of liver tissue provides direct visualisation of architectural changes in the liver parenchyma (the functional tissue of an organ, as opposed to the supportive or connective tissue). This yields diagnostic information not only on fibrosis, but also on many other histological features, including necroinflammation (a process involving cell death and tissue inflammation), steatosis (fatty liver disease), and hepatic deposits of iron or copper.

Percutaneous and transjugular liver biopsies are the two most common methods for obtaining a liver tissue sample. Both can be performed as a day procedure, and they do not routinely require hospitalisation. Laparoscopic biopsies and biopsies during laparotomy are other biopsy methods, but they are more invasive and are not routinely used.

Percutaneous liver biopsy, which involves inserting a biopsy needle through the skin and into the liver, lasts just a few seconds. It is performed under local anaesthesia, usually while the person holds their breath after expiration. Percutaneous liver biopsy is accompanied by several crucial drawbacks, including variable accessibility, high cost, sampling errors (or sampling variability), and inaccuracy due to inter- and intra-observer variability of pathological interpretations (Sebastiani 2006). Intra- and inter-observer variability of histopathological examination depends on the biopsy size and number of portal tracts (areas of tissue that contain specific structures examined for diagnostic purposes). Although some histopathologists consider a biopsy sample length of 20 mm and a minimum of 11 portal tracts adequate (Colloredo 2003), others have shown that even biopsies of 25 mm length can accurately diagnose and stage only 75% of people with liver fibrosis in hepatitis C (Bedossa 2003). However, as 25 mm or even 20 mm of liver tissue would probably require more than one pass with the biopsy needle – hence increasing the risk of adverse effects – a conservative approach that obtains a 15 mm sample and six portal tracts is still considered relatively reliable (Cholongitas 2006). Major and minor complications occur in up to 6% of people undergoing percutaneous liver biopsy. About 0.04% to 0.11% of these complications can be life-threatening and are related to technical factors, including the experience of operators, use of larger needles, needing more than one pass, and possibly not using ultrasonography before or during liver biopsy. The most important complication is bleeding following the passage of the biopsy needle, which is clinically significant in 1.1% to 1.6% of people (Bravo 2001).

Transjugular liver biopsy involves inserting a catheter through the jugular vein in the neck and into the hepatic vein in the liver. A biopsy needle is then passed through the catheter to obtain liver tissue. This alternative method is favoured when obtaining a sample from people defined as high-risk; that is, people with extreme obesity, gross ascites (the abnormal accumulation of fluid in the abdomen), severe coagulopathy (where the blood's ability to form clots is impaired), or in the case of previous failure of percutaneous liver biopsy. Transjugular liver biopsy has the advantage that the fibrous layer of connective tissue that surrounds the liver (the Glisson capsule) is not breached (except as a procedural complication from within the liver), and therefore bleeding complications are extremely rare (Kalambokis 2007).

In recent years, interest in identifying and describing liver fibrosis by using non-invasive surrogate markers has increased. Serum markers of liver fibrosis offer an attractive, cost-effective alternative to liver biopsy. In addition to being less invasive, they have practically no complications, little or no sampling errors, and small observer-related variability (Grigorescu 2006). Moreover, they can be used to prioritise people for chronic hepatitis C treatment based on disease stage, as well as for monitoring treatment response and fibrosis regression (Recommendation A1, EASL-ALEH 2015). However, none are liver-specific and can be affected by changes in excretion and clearance of the individual components of each test. Biological variables, such as age, sex, and food intake, can influence serum marker levels, particularly in the context of low-level fibrosis and direct serum markers (Lichtinghagen 2013).

Non-invasive tests of fibrosis should correctly classify people for a certain outcome in relation to the reference standard (true positive and true negative). False negative results occur when the test incorrectly rules out people with fibrosis – termed a 'missed diagnosis'. This leads to inappropriate delays in treatment initiation, surveillance for hepatocellular carcinoma, and endoscopic screening for varices, potentially affecting the prognosis of these high-risk individuals and causing serious harm. False positive results occur when the test incorrectly flags people without fibrosis. This leads to wrong prioritisation of antiviral treatment, and unnecessary surveillance of hepatocellular carcinoma or endoscopic screening for varices, potentially causing harm to these low-risk individuals due to anxiety and the side effects and risks that the treatments and surveillance investigations carry.

Target condition being diagnosed

Liver fibrosis is the excessive accumulation of extracellular matrix proteins, including collagen, that occurs in most types of chronic liver diseases. Liver fibrosis is assessed through liver histological scoring systems that use stages to describe and assess liver architecture and fibrosis. The METAVIR (Meta-analysis of Histological Data in Viral Hepatitis) scoring system stages fibrosis in five categories, from 0 to 4, with stage 0 (F0) signifying no fibrosis and stage 4 (F4), cirrhosis (METAVIR 1996). Stage 1 (F1) is characterised by portal tract fibrosis without septa formation (minimal scarring) and represents mild fibrosis. Stage 2 (F2) is characterised by portal tract fibrosis with infrequent septa formation (scarring around vessels within the liver) and represents significant fibrosis. Stage 3 (F3) is characterised by numerous septa but no cirrhosis and represents severe fibrosis.

We used the METAVIR classification system to define our diagnostic comparisons (METAVIR 1996): F0 to F1 compared to F2 to F4; F0 to F2 compared to F3 to F4; and F0 to F3 compared to F4.

Index test(s)

Several serum markers have been identified as potentially useful indicators of fibrosis, especially when combined with each other. Serum markers of fibrosis are commonly divided into direct and indirect markers. Direct markers are fragments of the liver matrix components produced by hepatic stellate cells during the process of extracellular matrix remodelling. Indirect markers include molecules released into the blood due to liver inflammation, molecules synthesised/regulated or excreted by the liver, and markers of processes commonly disrupted due to liver function impairment.

Direct and indirect markers may be used alone, but are more commonly used in combination with each other to produce composite scores. The calculation of such scores can be relatively simple or can be based on complicated formulas (e.g. those underlying FibroTest/FibroSure) (Grigorescu 2006). Both the Fibrosis-4 Index (FIB-4) and the Forns index are composite scores derived from indirect serum markers for fibrosis.

The FIB-4 score is a non-invasive test that combines standard biochemical values (platelets (PLT), alanine aminotransferase (ALT), and aspartate aminotransferase (AST)) with age. The formula is: age ([yr] x AST [U/L])/((PLT [10^9/L]) x ($ALT^{1/2}$ [U/L])). It was originally developed for use in hepatitis C virus (HCV)/human immunodeficiency virus (HIV) co-infection (Sterling 2006). The FIB-4 score has validated cut-offs described in the original study by Sterling and colleagues: the low cut-off (1.45) to rule out people with at least severe fibrosis (≥ F3), and the high cut-off (3.25) to rule in people with at least severe fibrosis (≥ F3) (Sterling 2006).

Forns and colleagues first developed the Forns index in a cohort of untreated people with chronic hepatitis C (Forns 2002a). The formula is: (7.811-3.131.ln(PLT count) + 0.781.ln(gamma glutamyl-transferase (GGT)) + 3.467.ln(age) - 0.014.(cholesterol)). The cut-offs for these tests were defined and validated in the first study phase: the low cut-off (4.2) to rule out people with at least significant fibrosis (≥ F2), and the high cut-off (6.9) to rule in people with at least significant fibrosis (≥ F2).

There is a published protocol for a Cochrane diagnostic test accuracy (DTA) review that aims to compare several non-invasive tests for diagnosing severe hepatic fibrosis and cirrhosis in adults with chronic hepatitis C (Pavlov 2015a). There is also a completed Cochrane DTA review on transient elastography (a test which measures stiffening of the liver caused by scarring) in people with alcoholic liver disease (Pavlov 2015b).

Clinical pathway

According to EASL Clinical Practice Guidelines, people with detectable HCV antibodies should have HCV RNA determined by a sensitive molecular method (EASL 2020). The contribution of comorbid conditions to the progression of liver disease should be evaluated along with the severity of liver disease. Following this, the stage of fibrosis should be assessed. EASL guidelines suggest that fibrosis should initially be assessed by non-invasive methods, including the FIB-4 score (EASL 2020). The guidelines suggest that, due to their low cost and wide accessibility, the FIB-4 score and the Forns index are well suited to rule out severe fibrosis in low prevalence settings (EASL 2021). By contrast, liver biopsy should be reserved for people in whom there is diagnostic uncertainty (EASL 2020).

Non-invasive tests such as the FIB-4 score may be used prior to initiating DAA therapy since it is imperative to determine whether advanced fibrosis or cirrhosis is present in order to define the appropriate treatment duration (eight versus 12 weeks, respectively) and post-treatment follow-up (EASL 2020). Whilst not specifically recommended in either guideline, the Forns index is an alternative non-invasive test to make this determination. However, neither of these tests should be used to evaluate whether there

have been changes in fibrosis stage after sustained virological response (SVR), as their reliability is decreased after treatment with DAAs (EASL 2020).

We aim to assess the degree to which the FIB-4 score and Forns index can be used as triage or replacement tests for liver biopsy, based on their diagnostic accuracy, as illuminated in this review.

Prior test(s)

Hepatitis C virus genotype, levels of transaminases, and liver synthetic function should be assessed before the initiation of treatment. Liver fibrosis stage should be assessed after the diagnosis of chronic hepatitis C and prior to the initiation of treatment. The FIB-4 score or the Forns index test could potentially be one of the first tests that people undergo after diagnosis of chronic hepatitis C.

Role of index test(s)

The FIB-4 score and Forns index are non-invasive methods for assessing liver fibrosis and are used as triage or replacement tests for liver biopsies.

Alternative test(s)

In addition to the FIB-4 score and the Forns index, several serum markers have been identified as possible useful indicators of fibrosis, especially when combined with each other.

Five direct indices are protected by patents and are currently commercially available: the FibroTest in Europe (Biopredictive, Paris, France) or FibroSure in the USA (LabCorp, Burlington, NC, USA), the Fibrometers (BioLiveScale, Angers, France), the FibroSpect II (Promotheus Laboratory Inc., San Diego, CA, USA), the ELF (Enhanced Liver Fibrosis Test, Siemens), and the Hepascore (PathWest, University of Western Australia, Australia) (Castera 2009). Commonly used indirect markers are the aspartate aminotransferase (AST) / alanine aminotransferase (ALT) ratio, the AST to Platelet Ratio Index (APRI) score, and the Lok index (Castera 2009).

There are also indirect markers that combine commonly available laboratory tests and epidemiological variables such as body mass index (BMI) or age, including the non-alcoholic fatty liver disease (NAFLD) fibrosis score, and BARD (BMI, AST/ALT Ratio, Diabetes) (Shah 2009).

Various imaging modalities may also be used to assist in the diagnosis of liver fibrosis. Transient elastography is based on elastometry and is performed with Fibroscan (Echosens, Paris, France). It is a non-invasive method, designed to measure liver stiffness (Castera 2008). Other forms of elastography include acoustic radiation force impulse (ARFI) and shear wave elastography (EASL 2021). In magnetic resonance elastography, low-frequency longitudinal mechanical waves are transmitted into the right lobe of the liver by a transducer placed against the lowest ribs at the back of a person in a supine position (Huwart 2008). Ultrasonography, computed tomography scan, or magnetic resonance imaging have traditionally been used to explore the liver. These methods are able to detect changes in the liver parenchyma when there is significant fibrosis (bridging fibrosis and mainly cirrhosis) and signs of portal hypertension (enlarged spleen, collateral venous circulation, enlarged portal vein). However, these methods are not particularly useful for identifying people with less advanced stages of fibrosis (Bonekamp 2009). None of the above-mentioned tests are part of the standard diagnostic work-up.

Rationale

The assessment of liver fibrosis severity in people with chronic hepatitis C is vital to determine the need for follow-up after successful antiviral treatment (EASL 2020). Liver biopsy is still considered the best reference standard for staging fibrosis. The advantages of liver biopsy are that it fulfils its purpose and that it provides information on concomitant liver diseases. However, non-invasive fibrosis tests have been developed as an alternative to liver biopsy in detecting and staging liver fibrosis and diagnosing cirrhosis (Nguyen 2011). According to the EASL Clinical Practice Guidelines on non-invasive tests for evaluation of liver disease, methods such as liver stiffness measurement or well-established combined biomarkers can be used instead of liver biopsy to assess liver disease severity prior to treatment at a safe level of predictability (EASL 2021).

Non-invasive tests have no known complications and infrequent sampling errors, and they also have the potential to be used in the staging of fibrosis, if they are accurate. If proven accurate, they could replace liver biopsy when the aetiological diagnosis of underlying liver disease is established and the only pertinent clinical question is the staging of fibrosis. This question is important, as the choice of a treatment regimen, post-treatment prognosis, and initiation of surveillance programmes for hepatocellular carcinoma and varices depends on the stage of fibrosis. Finding inexpensive, reliable methods of staging fibrosis is especially important in resource-poor settings, where access to more advanced diagnostic tests (such as transient elastography, MRI-elastography, or patented serum non-invasive tests) is limited.

However, there is no established algorithm regarding the use of non-invasive fibrosis tests in people with chronic hepatitis C. Cirrhosis is the only stage of liver disease that heralds specific screening strategies for varices and hepatocellular carcinoma; namely, endoscopic surveillance with upper gastrointestinal endoscopy and six-monthly ultrasound scan. There is no evidence from randomised clinical trials to estimate the potential risks of missing or falsely diagnosing fibrosis stages other than cirrhosis by using the index tests described above. Therefore, it is difficult to estimate acceptable performance diagnostic test accuracy for such stages. One method to estimate diagnostic test accuracy is to calculate the area under a receiver operating characteristic curve (AUROC). A ROC curve is a plot of sensitivity versus 1 − specificity. The AUROC is a measure of accuracy (Hanley 1982). An arbitrary threshold of 80% is used to define the acceptable limit of accuracy, with an AUROC below 0.8 generally considered too poor to be of value in clinical practice (EASL 2021). Whilst AUROCs are useful when estimating a ROC curve, we have performed a meta-analysis on fixed cut-offs, and therefore AUROCs do not form part of our review.

Currently, there is no Cochrane review assessing the diagnostic accuracy of the FIB-4 score or Forns index in chronic hepatitis C. There are several non-Cochrane systematic reviews on other non-invasive methods of fibrosis assessment (Chou 2013; Poynard 2011; Smith 2009; Xiao 2015). The present review aims to evaluate the diagnostic accuracy of the FIB-4 score and Forns index only in people with chronic hepatitis C, taking into consideration that

the accuracy of non-invasive tests is probably aetiology-dependent (Crossan 2015).

OBJECTIVES

To determine the diagnostic accuracy of the FIB-4 score and Forns index in staging liver fibrosis in people with chronic hepatitis C virus, using liver biopsy as the reference standard.

Secondary objectives

To compare the diagnostic accuracy of these tests for staging liver fibrosis in people with chronic hepatitis C, and to explore potential sources of heterogeneity:

- serum levels of alanine aminotransferase activity (Dufour 2000);
- people with and without HIV co-infection;
- studies at high risk of bias compared to studies at low or unclear risk of bias, applying the QUADAS-2 tool (Whiting 2011).

METHODS

Criteria for considering studies for this review

Types of studies

We included cross-sectional and case-control studies that evaluated the diagnostic accuracy of the FIB-4 score or Forns index for staging hepatic fibrosis, with liver biopsy as the reference standard, in participants with chronic hepatitis C. We imposed no restrictions on language or publication status, or whether data were collected prospectively or retrospectively. We considered conference abstracts and full publications. We excluded studies if the elapsed time between the index test and liver biopsy exceeded six months.

Participants

We included studies in adults diagnosed with chronic hepatitis C, irrespective of baseline characteristics, sex, or ethnicity, who had a FIB-4 score or Forns index score and liver biopsy. We excluded studies in which participants had causes of liver disease other than hepatitis C (such as alcohol abuse, hepatitis B co-infection, and schistosomiasis), or in which participants had a non-invasive fibrosis assessment after achieving sustained virological response to antiviral treatment for hepatitis C (i.e. undetectable viral loads in the blood 12 to 24 weeks after completing treatment) (EASL 2021). We included studies in people living with HIV.

Index tests

The two index tests of interest in this review were the FIB-4 score and Forns index. These are both indirect, non-invasive markers of liver fibrosis based on laboratory values. The FIB-4 score is calculated using alanine aminotransferase (ALT), aspartate aminotransferase (AST), platelets, and age. The Forns index is calculated using age, gamma glutamyl-transferase (GGT), platelets, and cholesterol. Both tests can be calculated according to the formulas described in Sterling 2006 and Forns 2002a, respectively. Sterling and colleagues developed the FIB-4 score specifically for diagnosing severe fibrosis (F3 and above), whilst Forns and colleagues developed their index for diagnosing significant fibrosis (F2 and above).

Target conditions

The target conditions were the presence of different stages of hepatic fibrosis in the selected population. The stages of hepatic fibrosis are widely defined according to the METAVIR histopathological scoring system (METAVIR 1996), where:

- F0 = absence of fibrosis;
- F1 = mild fibrosis;
- F2 = significant fibrosis;
- F3 = severe fibrosis; and
- F4 = cirrhosis.

The three target conditions we considered in this review, as defined by METAVIR stage, were:

- participants with at least significant fibrosis (stages F2 to F4) compared to participants with no significant fibrosis (stages F0 and F1);
- participants with at least severe fibrosis (stages F3 and F4) compared to participants without severe fibrosis (stages F0 to F2);
- participants with cirrhosis (stage F4) compared to participants with no cirrhosis (stages F0 to F3).

Reference standards

The reference standard used was histopathological examination of liver tissue (liver biopsy) obtained through percutaneous, transjugular, or laparoscopic biopsy (Cholongitas 2006).

Whilst liver biopsy is currently the only existing reference standard for diagnosing hepatic fibrosis stages in chronic hepatitis C infection, it can be scored using various systems such as the METAVIR system described above. Other scoring systems include those developed by Ishak 1995, Knodell 1981, and Scheuer 1991. If studies reported stage of fibrosis using a system other than METAVIR, we used a conversion grid adapted from Goodman 2007 to convert their scores to METAVIR scores (see Table 1).

Search methods for identification of studies

Electronic searches

We searched the Cochrane Hepato-Biliary Group Diagnostic Test Accuracy Studies Register, MEDLINE Ovid, Embase Ovid, Science Citation Index Expanded (Web of Science), CINAHL (Cumulative Index to Nursing and Allied Health Literature; EBSCOhost), and LILACS (Latin American and Caribbean Health Science Information database; Bireme), without language restrictions, from January 2003 to 13 April 2022.

The search terms used are described in the review's protocol (Kalafateli 2015), and are reported in Appendix 1.

Searching other resources

Our searches of the electronic databases listed above encompassed conference proceedings, which we considered for inclusion. We did not perform handsearches, as there is little evidence of their fruitfulness for obtaining reports of diagnostic test accuracy studies (Glanville 2010). We assessed the reference lists of all included studies, as well as other systematic reviews in this field, for further eligible titles. We repeated this process until no new titles were found (Greenhalgh 2005). We attempted to contact the

corresponding authors of papers on apparently eligible studies with incomplete published data, to try to obtain the missing data. A reminder email was sent two weeks after our initial contact. If we received no reply, we excluded the study.

Data collection and analysis

We followed the guidelines provided in the *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy* (Deeks 2023).

Selection of studies

After excluding duplicates, two review authors (MH/MZ or MH/AL) independently screened abstracts. Full texts were retrieved and assessed in the same manner. A third review author (ET) resolved any disagreements about eligibility, resulting in the final list of included studies.

We did not plan to use a methodological search filter due to the risk of excluding relevant search results. To reduce the risk of reporting bias, we included eligible studies published in languages other than English and conference proceedings. We did not test for publication bias due to the lack of validated methods for diagnostic test accuracy reviews.

Data extraction and management

Two review authors (MH/MZ or MH/AL) independently performed data extraction using a predefined electronic data extraction sheet. A third review author (ET) arbitrated any disagreements.

Where available, we reported the following details for each included study: country of origin, study design, temporality of data collection (i.e. retrospective/prospective), participant sampling method, type of publication, total and included study participants, epidemiological and laboratory characteristics, cut-offs used for the FIB-4 score and Forns index, stage of liver fibrosis assessed, histological scale used, and information for the QUADAS-2 risk of bias evaluation. Regarding sampling method, if participants were selected based on suspicion of liver fibrosis (i.e. a cohort of people known to have hepatitis C virus from a particular centre), we labelled this 'cohort-based sampling'; if participants were selected as either cases or controls, we labelled this 'case-control sampling' (Mathes 2019).

The minimal data requirement, for studies to be included in the review, was to provide sufficient data to calculate the true positive (TP), false positive (FP), true negative (TN), and false negative (FN) diagnostic values of the serological test, compared to the reference standard, liver biopsy, and the cut-off used for each condition. When any of these values were missing, we contacted the authors via email, as described in Searching other resources.

Assessment of methodological quality

Methodological quality of individual studies

Working independently, two review authors (MH/MZ or MH/AL) assessed the risk of bias and applicability concerns of included studies, using the QUADAS-2 tool (Whiting 2011). A third review author (ET) acted as arbitrator in case of disagreement. We contacted the study authors when data in the published report were not available or unclear. In such cases, we disregarded some of the data presented in the publication and used the data provided by the study authors through personal communication.

The QUADAS-2 tool evaluates the risk of bias in a study across four domains: participant selection, index test, reference standard, and flow and timing (Whiting 2011). The signalling questions and answers are summarised in Appendix 2. Our QUADAS-2 assessments are summarised in Figure 1, Figure 2, and Figure 3.

Figure 1. Risk of bias and applicability concerns graph: review authors' judgements about each domain presented as percentages across included studies

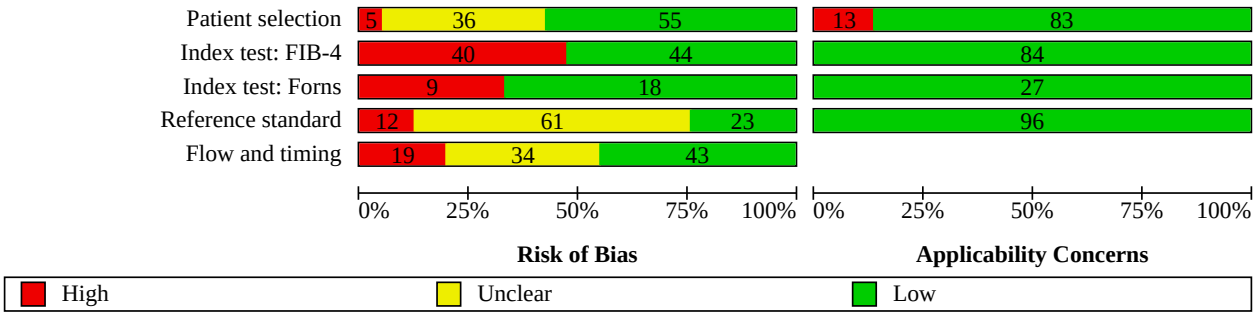


Figure 2. Risk of bias and applicability concerns summary: review authors' judgements about each domain for each included study testing the FIB-4 score.

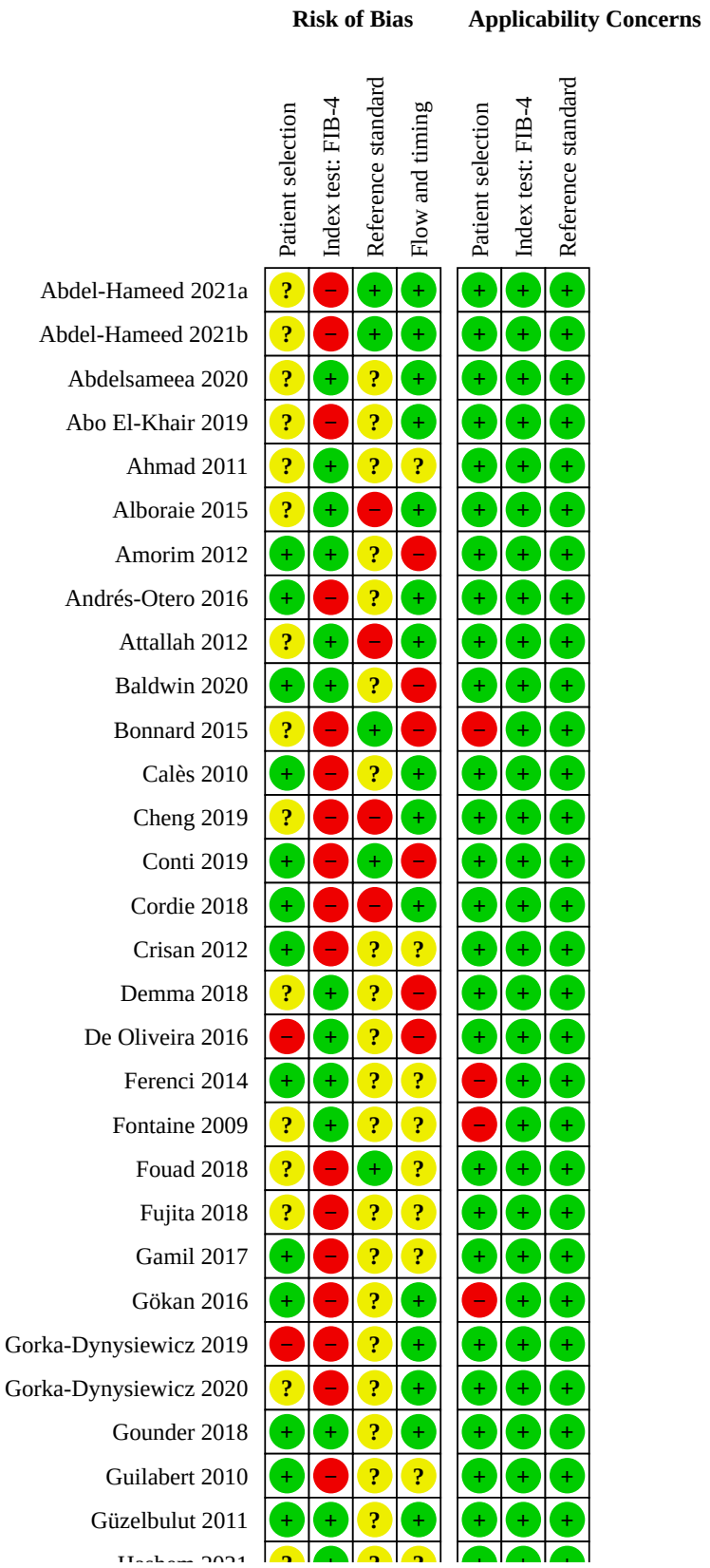


Figure 2. (Continued)

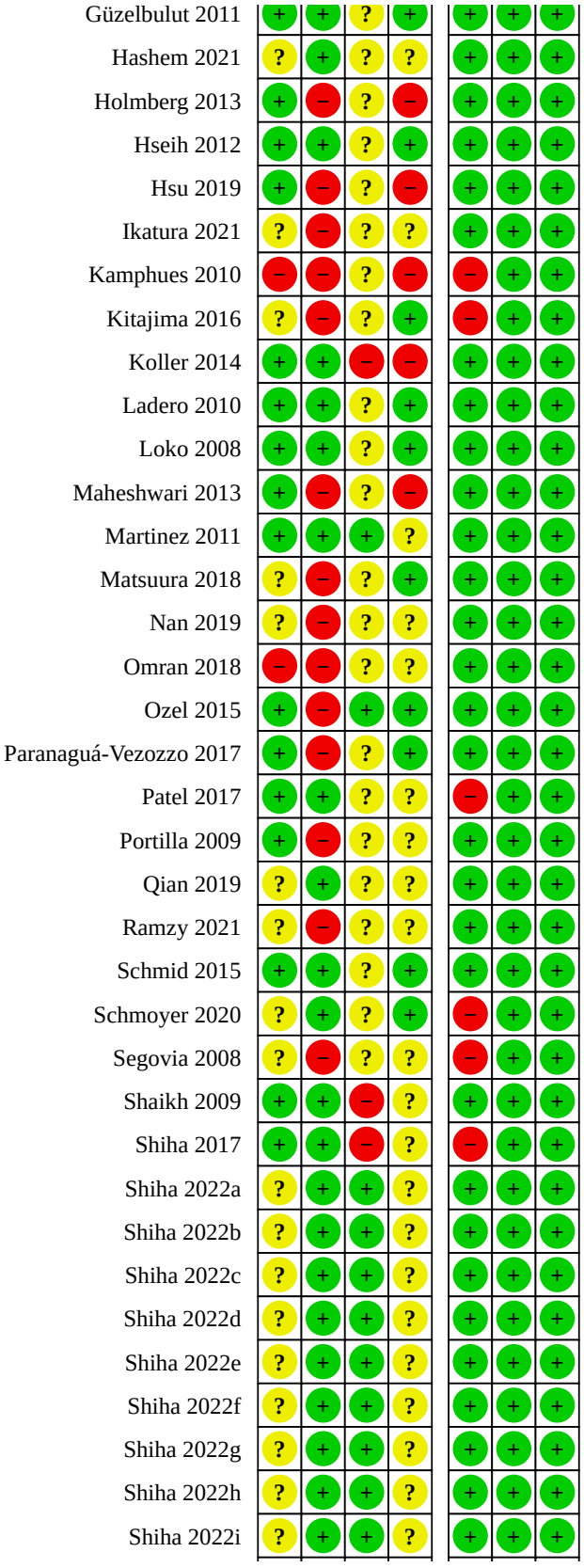


Figure 2. (Continued)

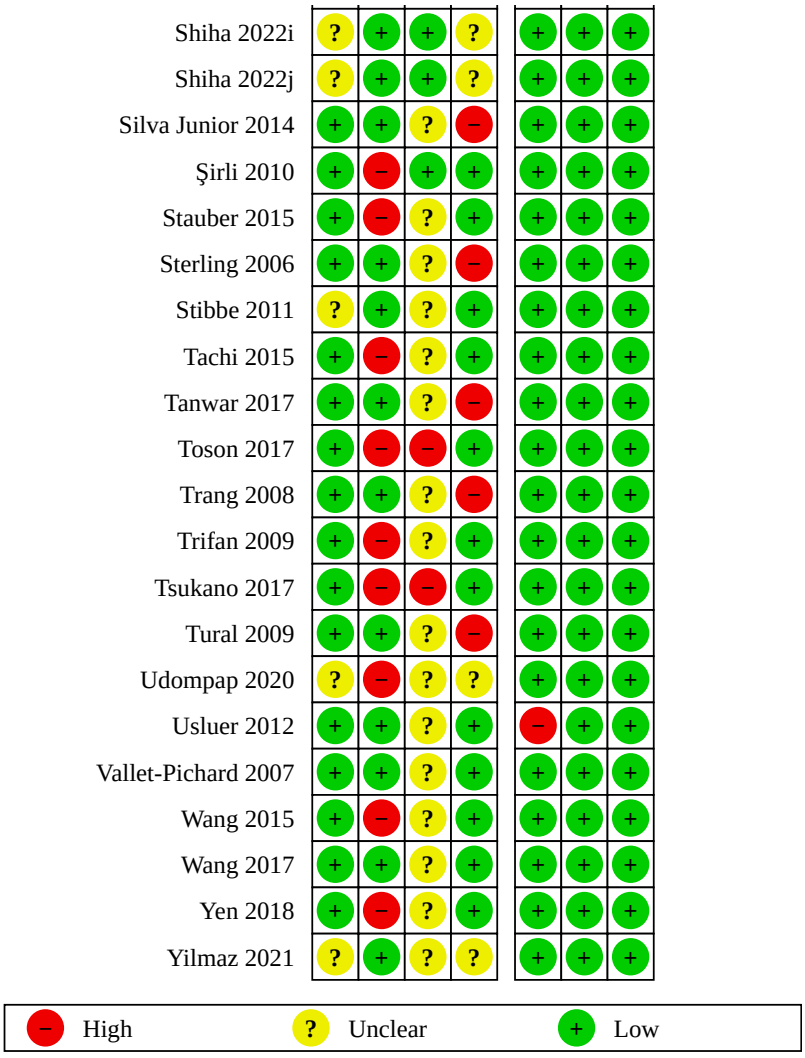
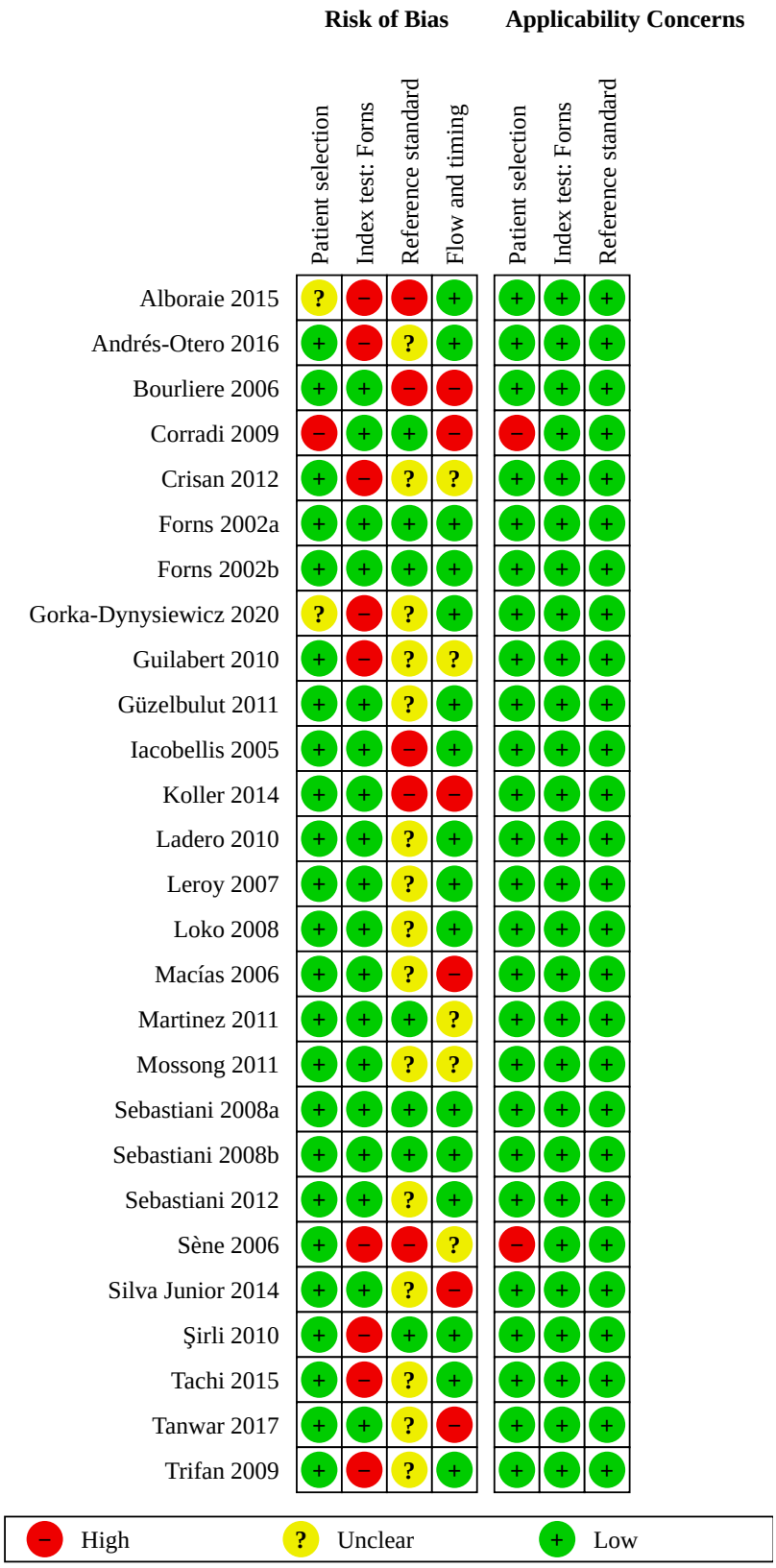


Figure 3. Risk of bias and applicability concerns summary: review authors' judgements about each domain for each included study using the Forns index.



We deemed studies to have a low risk of bias for participant selection if they enroled consecutive or random participants, avoided case-control design, and if we judged there to have been no inappropriate exclusions. We deemed studies to have a low risk of bias for the index test if the assessors interpreted the index tests without knowledge of biopsy results, and if they used a prespecified threshold. We deemed studies to have a low risk of bias for the reference standard if authors reported a requirement of at least six portal tracts in the biopsy specimen for it to be deemed valid, and if authors stated that the pathologist was unaware of the index test result. We judged studies to be at low risk of bias for flow and timing if they included all participants who received the index test in the analysis, and if the time between the biopsy and the index test was three months or less. We considered studies in which the elapsed time between index test and biopsy was three to six months to be at high risk of bias in this domain. We excluded studies in which the elapsed time between index test and biopsy exceeded six months, as mentioned in Types of studies. We adopted this six-month threshold because this is the minimum time interval in which a change in fibrosis stage is thought to occur, and is commonly used in clinical trials to observe changes in histological stage (Francque 2021). For all four domains, if key information was unavailable, we judged studies to have an unclear risk of bias.

We judged studies to have an overall low risk of bias if we judged there was a low risk of bias across all four domains. If there was a high risk of bias in any single domain, we deemed the study's overall risk of bias to be high. If there was an unclear risk of bias in any single domain, we deemed the study's overall risk of bias to be unclear.

Quality of the overall body of evidence

We assessed the overall quality of the body of evidence according to the GRADE guidelines (Guyatt 2011), by examining the evidence against four domains: risk of bias, imprecision, inconsistency, and indirectness. We did not evaluate overall publication bias in the body of evidence as there are no validated methods for doing so in diagnostic test accuracy reviews. We assessed the risk of bias for individual studies using the QUADAS-2 tool, as described above, and the overall risk of bias in the included studies by judging the overall spread of high-, unclear-, and low-risk studies. We assessed imprecision by examining the spread of the confidence intervals (CIs) in the meta-analysis summary estimates and judging whether clinical action would differ if the upper versus the lower boundary of the CI represented the truth. We assessed inconsistency by inspecting the forest plots of the included studies and examining the similarity of point estimates between studies. We assessed indirectness by considering any differences between the populations, setting, and index tests in the included studies compared to our review question.

We created a summary of findings table to present the review's most salient findings. This table includes the results of our assessment of the overall quality of the body of evidence within the review, and therefore the certainty of the results, developed in accordance with the GRADE framework (GRADEpro GDT).

Statistical analysis and data synthesis

We carried out all analyses according to the guidelines provided in the *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy* (Deeks 2023).

We treated histological stages as dichotomous variables; that is, lower than versus equal to or higher than a specific histological stage. The three target conditions we considered in this review, as defined by METAVIR stage, were:

- participants with at least significant fibrosis (stages F2 to F4) compared to participants with no significant fibrosis (stages F0 and F1);
- participants with at least severe fibrosis (stages F3 and F4) compared to participants without severe fibrosis (stages F0 to F2);
- participants with cirrhosis (stage F4) compared to participants with no cirrhosis (stages F0 to F3).

For each included study, we created two-by-two tables of the number of TPs, TNs, FPs, and FNs for each index test (FIB-4 score or Forns index) and each of the three target conditions (significant fibrosis, severe fibrosis, and cirrhosis). We report on the diagnostic accuracy of each of the low and high cut-offs. These are stand-alone cut-offs that can be used either to rule out the presence of a certain stage of fibrosis with high sensitivity or to rule in the presence of a fibrosis stage with high specificity. We calculated the sensitivity and specificity with their 95% confidence intervals (CIs) for each test in each target condition.

We recorded the cut-offs used in each study. We expected the cut-offs used in the included studies to vary, despite the fact that specific cut-offs for both the FIB-4 score and Forns index were developed and validated in the original publications (Forns 2002a; Sterling 2006). The original validated cut-offs are:

- FIB-4 score: 1.45 to rule out the presence of severe fibrosis (≥ F3); 3.25 to rule it in;
- Forns index: 4.2 to rule out the presence of significant fibrosis (≥ F2); 6.9 to rule it in.

In order to avoid the limits this heterogeneity in cut-off values poses to a summary estimate, and to provide meaningful results for clinical practice, we only included studies in the meta-analysis if the reported cut-offs were in a narrow range around the original validated cut-offs (+/- 0.15 for FIB-4; +/- 0.3 for Forns index). The rationale for this is that the use of diagnostic tests for clinical decision-making requires specific threshold values, and if we had grouped together vastly different cut-offs, it would provide misleading information that would not be useable in clinical practice. For the FIB-4 score, this meant we only included in the meta-analysis those studies with a low cut-off between 1.30 and 1.60 and a high cut-off between 3.10 and 3.40. For the Forns index, this meant we only included in the meta-analysis those studies with a low cut-off between 3.9 and 4.5 and a high cut-off between 6.6 and 7.2.

For each combination index test/target condition/cut-off value, we performed a descriptive analysis of the included studies by reporting separate forest plots for sensitivity and specificity. We also plotted studies on the receiver operating characteristic (ROC) space (sensitivity against 1 − specificity). To facilitate interpretation, we represented these results graphically as sensitivity against specificity on a reversed scale (from 1 to 0, rather than from 0 to 1). We performed meta-analyses using the bivariate model, and we obtained estimates of summary sensitivity, specificity, positive and negative likelihood ratios (LR+ and LR-, respectively),

with their 95% CIs from the estimates of the bivariate model's parameters. We used pooled likelihood ratios arising from the meta-analysis to calculate post-test probabilities, starting from some pre-test probabilities. We used the median prevalence of the included studies as an estimate of the pre-test probability for each target condition and test. When computing median prevalence, we excluded case-control studies.

For the purposes of our meta-analyses, we treated the results from the index tests as dichotomous variables (either above or below a specified threshold). However, in the clinical environment, it is possible to consider both the rule-in and rule-out thresholds at the same time for an individual's result, in which case there are three outcomes: below the low cut-off (diagnosis ruled out), above the high cut-off (diagnosis ruled in), or indeterminate (in between the cut-offs). For each index test/target condition combination, we calculated the number of people who would have received an indeterminate test: the so-called "grey area". This represents the number of people whose test results fell above the low/rule-out cut-off, but below the high/rule-in cut-off, and would therefore have to undergo further testing. To calculate this grey area, we considered only studies reporting both the high and low cut-offs when applying the index test to the target condition. The aim of reporting these data is mainly descriptive, in order to show in how many people a decision is not possible when using both cut-offs at the same time.

We made pair-wise comparisons between tests by adding an index test covariate to the bivariate model. We assessed the significance of differences in test accuracy by using the log-likelihood ratio test for comparison of models with and without the index test covariate term. We calculated relative sensitivity and specificity, with their 95% CIs. We performed only indirect comparisons, as we had insufficient data for direct comparisons.

For the comparisons, we considered two-sided P values of less than 0.05 as statistically significant. We performed all the statistical analyses using SAS statistical software, release 9.4 (SAS Institute Inc., Cary, NC, USA), and macro METADAS (Deeks 2023).

Investigations of heterogeneity

We investigated sources of heterogeneity by adding covariates to the bivariate model. The covariates we used were: serum levels of alanine transferase activity (ALT), different levels of inflammation according to liver biopsy, different quality of liver biopsy samples, and people with and without HIV co-infection. We assessed the statistical significance of the covariate effect by using the log-likelihood ratio test for comparison of models with and without the covariate term.

We chose ALT as a potential source of heterogeneity because it is a component in the FIB-4 score, but has low specificity for hepatitis C virus-caused fibrosis, since ALT levels may be raised in multiple other conditions, including alcohol-use disorder or obesity. Therefore, the index test may perform differently in each subset. For this analysis, we used an arbitrary threshold of two times the upper limit of normal: we compared studies with mean ALT values greater than 80 IU/L to studies with a mean ALT less than or equal to 80 IU/L.

We chose HIV as a potential source of heterogeneity since both the pathophysiology of the disease process itself and the administration of antiretroviral therapy are associated with hepatic steatosis and subsequent derangement of the liver laboratory variables that are used to calculate both the FIB-4 score and Forns index. Therefore, the index test may perform differently in each subset. For this analysis, we compared studies including one or more participants living with HIV to studies including no participants living with HIV.

We chose the level of histological inflammation as a potential source of heterogeneity because this independent cause for raised liver laboratory values may impact on the diagnostic accuracy of the FIB-4 score and Forns index, which rely on these values. For this analysis, we compared studies with a hepatitis activity index (HAI) of 0, 1, or 2 to studies with a mean HAI of 3 and 4.

Finally, we selected the quality of liver biopsy specimens as a potential source of heterogeneity because studies that included low-quality specimens as the reference standard may obtain different results for the diagnostic accuracy of the index tests. For this analysis, we compared studies that restricted biopsy samples to a minimum of six portal tracts to studies that included biopsy samples with a specified lower minimum number of portal tracts.

Sensitivity analyses

We performed a sensitivity analysis by excluding studies at high or unclear risk of bias and assessing the influence, or lack thereof, on the results. As noted in Assessment of methodological quality, we classified a study as having a high or unclear risk of bias overall if we judged it to have a high or unclear risk of bias in at least one of the domains of the QUADAS-2 tool (Whiting 2011).

We performed a further sensitivity analysis by excluding studies at high or unclear risk of bias in each of the four domains assessed by the QUADAS-2 tool (selection of study participants, index test, reference standard, and flow and timing).

Finally, to see if publication bias had an impact on the results, we conducted a sensitivity analysis by excluding studies published in abstract form only.

Assessment of reporting bias

We did not assess publication bias due to the lack of validated methods for diagnostic test accuracy reviews.

RESULTS

Results of the search

A total of 17,517 references were identified and screened from searches performed on 13 April 2022, summarised in Figure 4. The following databases were searched: the Cochrane Hepato-Biliary Diagnostic Test Accuracy Studies Register (n = 42), MEDLINE Ovid (n = 6045), Embase Ovid (n = 8074), Science Citation Index Expanded (ISI Web of Science) (n = 2412), CINAHL (EBSCOhost) (n = 79), and LILACS (Bireme) (n = 865). Four additional references were retrieved from the reference lists of included studies. After exclusion of 11,052 duplicates, 6469 records remained for screening. Of these, we retrieved 206 full texts, and identified 76 studies eligible for inclusion in the review.

Figure 4. PRISMA flow diagram detailing database searches, number of abstracts screened, and number of full texts reviewed. Date of search: 13 April 2022

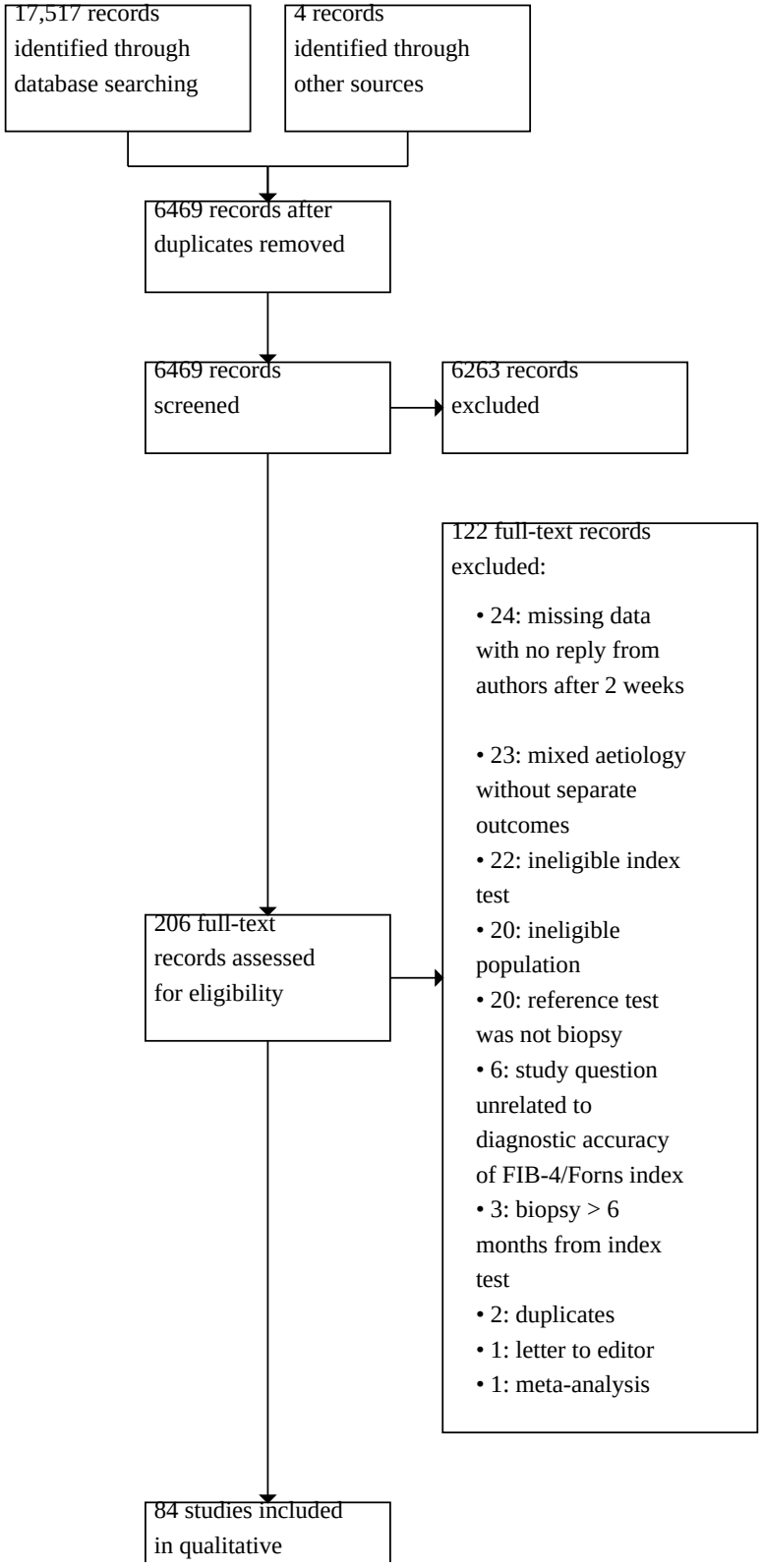
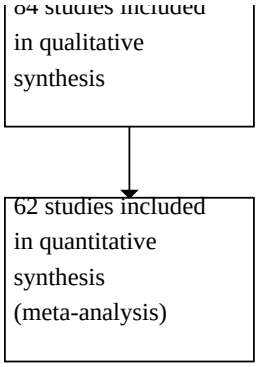


Figure 4. (Continued)



We identified a further 32 potentially eligible studies, with insufficient data reported for the purposes of our review (for example, specific sensitivity and specificities not reported). We emailed the corresponding authors of these studies, and received replies from 17. In eight instances, the authors provided us with the missing data, and we were able to include these studies in our review (Abdel-Hameed 2021a; Abdelsameea 2020; Ferenci 2014; Hsu 2019; Ikatura 2021; Martinez 2011; Shiha 2022a; Sterling 2006). We excluded the 24 remaining studies after receiving no reply from the corresponding authors to the two-week reminder email requesting additional data.

Thus, we included a total of 84 studies with 107,583 participants in the review (Figure 4). Eighty-two studies were reported in English, and two were reported in Spanish (Guilabert 2010; Portilla 2009).

Included studies

The 84 studies enroled participants from a total of 28 countries. The median year of publication was 2015, with a range from 2002 to 2021. The country producing the largest number of studies was Egypt (n = 13, 15%), followed by the USA (n = 9, 11%), Spain (n = 8, 10%), France (n = 7, 8%), and Japan (n = 6, 7%). There were no studies from South America, Oceania or sub-Saharan Africa. There were two reported case-control diagnostic accuracy studies (Corradi 2009; Gorka-Dynysiewicz 2019). The remaining 82 studies were designed as cross-sectional diagnostic accuracy studies sampling patients from prespecified HCV cohorts. Twenty-five studies (30%) were multicentre studies; 59 studies (70%) were single centre studies. All included centres were referral centres; no studies were carried out in primary care. The median disease prevalence across all 84 included studies, according to biopsy, were: 46.2% of participants with no to mild fibrosis (F0, F1), 32.2% with significant fibrosis (F2), 11.5% with severe fibrosis (F3), and 10.2% with cirrhosis (F4).

Seventy-four studies with a total of 103,560 participants reported the FIB-4 score. Twenty-five studies with a total of 6980 participants reported the Forns index. Fifteen studies reported both the FIB-4 score and Forns index (Alboraie 2015; Andrés-Otero 2016; Crisan 2012; Guilabert 2010; Gorka-Dynysiewicz 2020; Güzelbulut 2011; Koller 2014; Ladero 2010; Loko 2008; Martinez 2011; Silva Junior 2014; Şirli 2010; Tachi 2015; Tanwar 2017; Trifan 2009). Fifty-nine studies reported solely on the FIB-4 score and 10 studies reported solely on the Forns index.

As we anticipated (see Statistical analysis and data synthesis), the studies employed a large range of different cut-offs. Twenty-two studies used cut-offs that fell outside our prespecified narrow ranges around the original validated cut-offs (i.e. +/- 0.15 for the FIB-4 score; +/- 0.3 for the Forns index). Thus, ultimately, we included 62 studies (73%) with 100,605 participants in the meta-analyses.

Four studies reported the performance of the index test in more than one distinct cohort of participants (Abdel-Hameed 2021a; Forns 2002a; Sebastiani 2008a; Shiha 2022a). Since the reported sensitivities and specificities were unique to each cohort, we created distinct 2x2 tables of the TP/FP/TN/FN values for each, and treated each cohort as a unique study for the purposes of the meta-analysis. For this reason, the cohorts from these studies are found under separate references with suffixes a, b, c, and so on in the Characteristics of included studies tables.

Methodological quality of included studies

Methodological quality of individual studies

We provide detailed quality assessments of the included studies in Characteristics of included studies, and summarise this information in Figure 1, Figure 2, and Figure 3.

Overall, we judged only two studies to have an overall low risk of bias (i.e. low risk of bias across all domains) (Forns 2002a; Sebastiani 2008a). Twenty-three studies (27%) had an overall unclear risk of bias, and the remaining 59 studies (70%) were deemed to have an overall high risk of bias.

Overall, we judged 13 of 84 included studies (15%) to be of high concern for applicability. Of the 74 studies that reported on the FIB-4 score, we deemed 11 (15%) to have high concerns for applicability. Of the 25 studies that reported on Forns index, we deemed two (8%) to have high concerns for applicability.

None of the 74 studies that reported on the FIB-4 score had an overall low risk of bias. There was an overall high risk of bias in 53 studies (72%), and an overall unclear risk of bias in 21 studies (28%). We judged the participant selection process to be high risk

in four studies (5%), low risk in 44 studies (59%), and unclear risk in 26 studies (35%). We judged the conduct and interpretation of the FIB-4 score to be high risk in 39 studies (53%) and low risk in the remaining 35 studies (47%). We rated the conduct and interpretation of liver biopsy to be high risk in nine studies (12%), low risk in eight studies (11%), and unclear risk in 57 studies (77%). We assessed the flow and timing of the studies that reported on the FIB-4 score to be high risk in 16 studies (22%), low risk in 35 studies (47%), and unclear risk in 23 studies (31%).

Two of the 25 studies that reported on the Forns index had an overall low risk of bias (8%). There was an overall high risk of bias in 16 studies (64%), and an overall unclear risk of bias in seven studies (28%). We judged the participant selection process to be high risk in one study (4%), low risk in 22 studies (88%), and unclear risk in two studies (8%). We rated the conduct and interpretation of the Forns index to be high risk in nine studies (36%) and low risk in the remaining 16 (64%) studies. We judged the conduct and interpretation of liver biopsy to be high risk in five studies (20%), low risk in five studies (20%), and unclear risk in 15 studies (60%). We assessed the flow and timing of the studies that reported on the Forns index to be high risk in six studies (24%), low risk in 14 studies (56%), and unclear risk in five studies (20%).

Of the five studies that we judged to be at high risk of bias regarding participant sampling, the most common reasons were due to a case-control design (Corradi 2009; Gorka-Dynysiewicz 2019), or for inappropriately excluding participants due to obesity (Kamphues 2010; Omran 2018). All 48 studies (100%) that we judged to be at high risk of bias regarding the index test were judged as such because they did not use a prespecified threshold for the index test. We judged 14 studies (100%) to be at high risk of bias regarding the reference standard due to inappropriate inclusion of biopsy specimens that contained fewer than six portal tracts. We judged 22 studies to be at high risk of bias regarding flow and timing: in 12 studies (55%), the elapsed time between index test and biopsy was three to six months; and in 10 studies (45%), participants who received both the index test and biopsy were not included in the analysis.

Forty-three of the included studies did not use prespecified thresholds when assessing the diagnostic accuracy of the index test but instead derived the thresholds 'a posteriori'. We therefore assessed these studies to be at high risk of bias with regard to the index test, due to the risk of overestimating the diagnostic accuracy. This was the single largest issue in the overall body of data in our review. It accounts for 43 of the 59 (73%) studies that had an overall high risk of bias and is therefore the leading reason we judged studies to have an overall high risk of bias. Twenty-one of these studies contributed data to the meta-analysis as the derived cut-offs fell within our acceptable cut-off ranges.

Quality of the overall body of evidence

We rated the overall body of evidence for the performance of the FIB-4 score in ruling out severe fibrosis or worse, using the low cut-off, as 'very low'. We downgraded the evidence by one level for risk of bias due to a high number of studies with an unclear risk of bias. We did not downgrade for imprecision as 95% CIs for summary estimates were narrow enough that clinical action would not differ if the upper versus the lower boundary of the CI represented the truth. We downgraded by two levels for inconsistency due to extreme dissimilarity between the point estimates of individual studies. We did not downgrade for indirectness as there were low applicability concerns.

We rated the overall body of evidence for the performance of the FIB-4 score in ruling in severe fibrosis or worse, using the high cut-off, as 'low'. We downgraded by one level for risk of bias due to a high number of studies with an unclear risk of bias. We did not downgrade for imprecision as 95% CIs for summary estimates were narrow enough that clinical action would not differ if the upper versus the lower boundary of the CI represented the truth. We downgraded by one level for inconsistency due to dissimilarity between the point estimates of individual studies. We did not downgrade for indirectness as there were low applicability concerns.

We rated the overall body of evidence for the performance of the Forns index in ruling out significant fibrosis or worse, using the low cut-off, as 'low'. We downgraded by one level for risk of bias due to a high number of studies with an unclear risk of bias. We did not downgrade for imprecision as 95% CIs for summary estimates were narrow enough that clinical action would not differ if the upper versus the lower boundary of the CI represented the truth. We downgraded by one level for inconsistency due to dissimilarity between the point estimates of individual studies. We did not downgrade for indirectness as there were low applicability concerns.

We rated the overall body of evidence for the performance of the Forns index in ruling in significant fibrosis or worse, using the high cut-off, as 'very low'. We downgraded by one level for risk of bias due to a high number of studies with an unclear risk of bias. We downgraded by one level for imprecision as 95% CIs for summary estimates were wide enough that clinical action would differ if the upper versus the lower boundary of the CI represented the truth (i.e. a true LR+ 27.2 would confer very different clinical significance compared to LR+ 5.7). We downgraded by one level for inconsistency due to dissimilarity between the point estimates of individual studies. We did not downgrade for indirectness as there were low applicability concerns.

Findings

We summarise our findings in Summary of findings 1. Table 2 shows pooled estimates for each test/target condition/cut-off combination. We report our findings in detail below, derived from the 62 studies (with 100,605 participants) that reported cut-offs in a narrow range around the original validated cut-offs (+/- 0.15 for FIB-4; +/- 0.3 for Forns index).

FIB-4

Significant fibrosis (≥ F2)

Low cut-off ~ 1.45

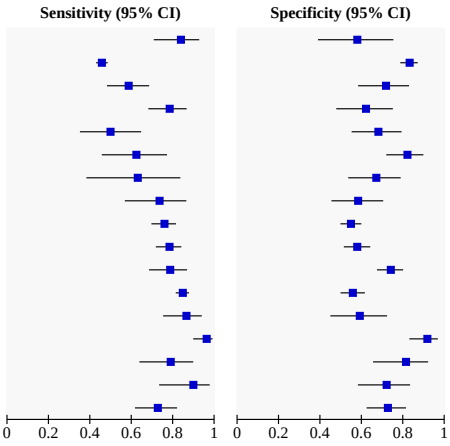
Seventeen study cohorts (from 17 studies) with 5098 participants provided data assessing the accuracy of the low cut-off of the FIB-4 score for diagnosing significant fibrosis or worse (≥ F2) (Amorim 2012; Andrés-Otero 2016; Conti 2019; De Oliveira 2016; Ferenci 2014; Fontaine 2009; Gounder 2018; Guilabert 2010; Ladero 2010; Ozel 2015; Paranaguá-Vezozzo 2017; Schmid 2015; Shaikh 2009; Shiha 2017; Trang 2008; Usluer 2012; Yilmaz 2021).

These studies used cut-off values ranging from 1.39 to 1.53, with 10 studies (59%) using the validated cut-off of 1.45. The sensitivity of

the low cut-off for the FIB-4 score for the diagnosis of F2 fibrosis or worse ranged from 46% to 96%, and the specificity ranged from 55% to 92% (Figure 5).

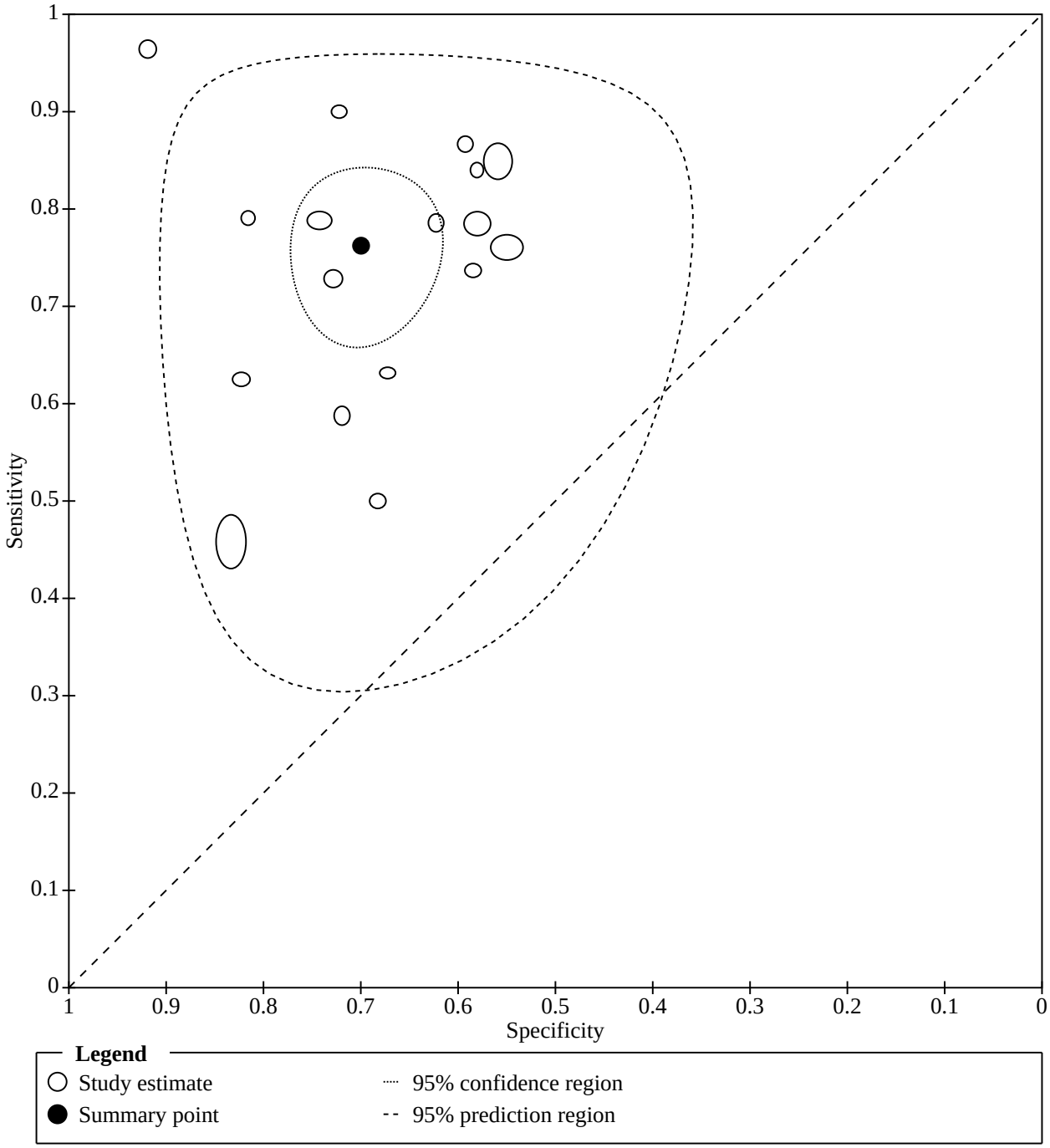
Figure 5. Forest plot of FIB-4 score for F2 – studies with cut-off around 1.45

Study	TP	FP	FN	TN	Cutoff	Sensitivity (95% CI)	Specificity (95% CI)
Trang 2008	42	13	8	18	1.39	0.84 [0.71, 0.93]	0.58 [0.39, 0.75]
Ferenci 2014	581	54	687	270	1.4	0.46 [0.43, 0.49]	0.83 [0.79, 0.87]
Guilabert 2010	57	16	40	41	1.4	0.59 [0.48, 0.69]	0.72 [0.58, 0.83]
Ozel 2015	66	20	18	33	1.4	0.79 [0.68, 0.87]	0.62 [0.48, 0.75]
Fontaine 2009	24	20	24	43	1.45	0.50 [0.35, 0.65]	0.68 [0.55, 0.79]
Amorim 2012	25	14	15	65	1.45	0.63 [0.46, 0.77]	0.82 [0.72, 0.90]
Usluer 2012	12	19	7	39	1.45	0.63 [0.38, 0.84]	0.67 [0.54, 0.79]
Schmid 2015	28	27	10	38	1.45	0.74 [0.57, 0.87]	0.58 [0.46, 0.71]
Shiha 2017	162	176	51	215	1.45	0.76 [0.70, 0.82]	0.55 [0.50, 0.60]
Ladero 2010	146	102	40	141	1.45	0.78 [0.72, 0.84]	0.58 [0.52, 0.64]
Gounder 2018	67	51	18	147	1.45	0.79 [0.69, 0.87]	0.74 [0.68, 0.80]
De Oliveira 2016	433	127	77	161	1.45	0.85 [0.81, 0.88]	0.56 [0.50, 0.62]
Yilmaz 2021	52	22	8	32	1.45	0.87 [0.75, 0.94]	0.59 [0.45, 0.72]
Shaikh 2009	81	6	3	68	1.45	0.96 [0.90, 0.99]	0.92 [0.83, 0.97]
Paranaguá-Vezozzo 2017	34	7	9	31	1.47	0.79 [0.64, 0.90]	0.82 [0.66, 0.92]
Andrés-Otero 2016	27	15	3	39	1.5	0.90 [0.73, 0.98]	0.72 [0.58, 0.84]
Conti 2019	59	25	22	67	1.53	0.73 [0.62, 0.82]	0.73 [0.63, 0.82]



We combined these data in a meta-analysis, and obtained the following summary estimates for the FIB-4 score using a low cut-off (~ 1.45) for the diagnosis of significant fibrosis (≥ F2): sensitivity 76.2% (95% CI 68.9% to 82.3%); specificity 70.0% (95% CI 64.0% to 75.4%); LR+ 2.5 (95% CI 2.1 to 3.1); LR- 0.34 (95% CI 0.25 to 0.45). Figure 6 shows a graphical representation of studies in the receiver operating characteristic (ROC) space.

Figure 6. Summary ROC plot of FIB-4 for F2 – studies with cut-off around 1.45 The circles represent individual studies.
The solid circle represents the summary estimate of sensitivity and specificity.
The dotted line represents the 95% confidence regions.
The dashed line represents the 95% prediction regions.



Amongst these study cohorts, the median prevalence of significant fibrosis was 46.8% (interquartile range (IQR) 36.9% to 61.7%). Using this value as a pre-test probability, we obtained a post-test probability of 68.7% (95% CI 64.9% to 73.2%) when the test was positive and a post-test probability of 23.0% (95% CI 18.0% to 28.4%) when the test was negative.

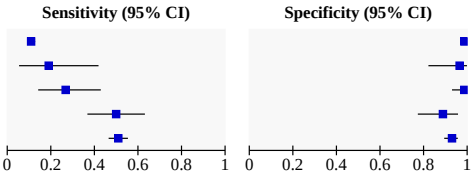
High cut-off ~ 3.25

Five study cohorts (from five studies) with 2673 participants provided data assessing the high cut-off for the FIB-4 score for the diagnosis of significant fibrosis or worse (≥ F2) (Amorim 2012; De Oliveira 2016; Ferenci 2014; Patel 2017; Yilmaz 2021).

The cut-off values used on these study cohorts ranged from 3.2 to 3.27. The validated cut-off of 3.25 was used on three occasions (60%). The sensitivity of the high cut-off for the FIB-4 score for the diagnosis of significant fibrosis or worse (≥ F2) ranged from 11% to 51%, and the specificity ranged from 88% to 99% (Figure 7).

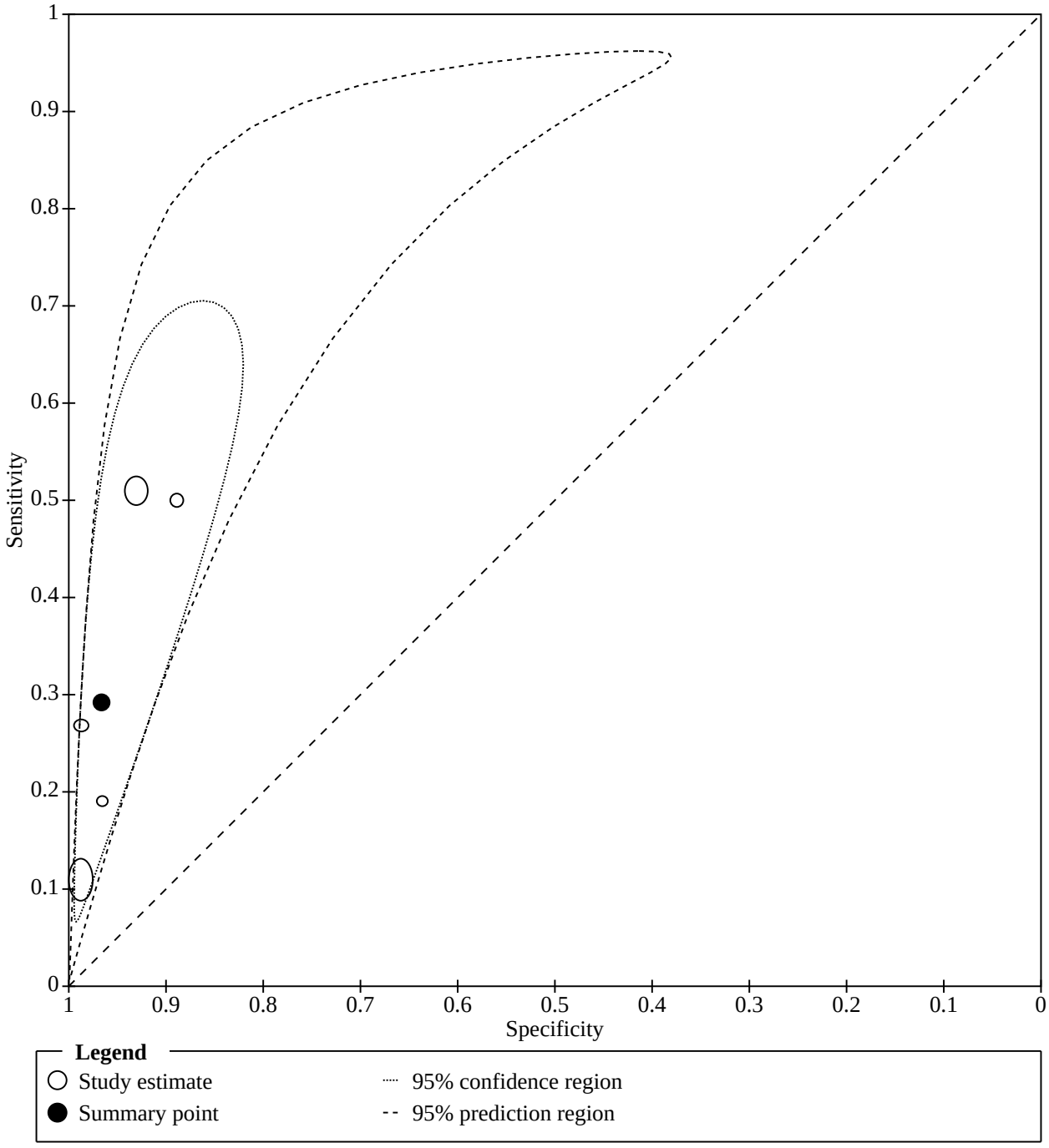
Figure 7. Forest plot of FIB4 for F2 - Studies with cut-off around 3.25.

Study	TP	FP	FN	TN	Cutoff	Sensitivity (95% CI)	Specificity (95% CI)
Ferenci 2014	139	4	1129	320	3.2	0.11 [0.09, 0.13]	0.99 [0.97, 1.00]
Patel 2017	4	1	17	28	3.25	0.19 [0.05, 0.42]	0.97 [0.82, 1.00]
Amorim 2012	11	1	30	77	3.25	0.27 [0.14, 0.43]	0.99 [0.93, 1.00]
Yilmaz 2021	30	6	30	48	3.25	0.50 [0.37, 0.63]	0.89 [0.77, 0.96]
De Oliveira 2016	260	20	250	268	3.27	0.51 [0.47, 0.55]	0.93 [0.89, 0.96]



Combining these data via meta-analysis yielded the following summary estimates for the FIB-4 score using a high cut-off (~ 3.25) for the diagnosis of significant fibrosis or worse (≥ F2): sensitivity 29.2% (95% CI 15.8% to 47.6%); specificity 96.6% (95% CI 92.6% to 98.5%); LR+ 8.7 (95% CI 5.5 to 13.7); LR- 0.73 (95% CI 0.60 to 0.90) (Figure 8).

Figure 8. Summary ROC plot of FIB4 for F2 - Studies with cut-off around 3.25. The circles represent individual studies.
The solid circle represents the summary estimate of sensitivity and specificity.
The dotted line represents the 95% confidence regions.
The dashed line represents the 95% prediction regions.



Amongst these study cohorts, the median prevalence of significant fibrosis was 52.6% (IQR 42.0% to 63.9%). Using this value as a pre-test probability, we obtained a post-test probability of 90.6% (95% CI 85.9% to 93.8%) when the test was positive and a post-test probability of 44.8% (95% CI 40.0% to 50.0%) when the test was negative.

Indeterminate results (grey area)

This represents the number of people for whom the test gives an inconclusive result (i.e. cannot rule in using high cut-off and cannot rule out using low cut-off). When calculating this grey area, we considered only the four studies reporting both the high and low cut-offs when applying the FIB-4 score to significant fibrosis (Amorim 2012; De Oliveira 2016; Ferenci 2014; Yilmaz 2021). If we consider all the 2623 participants enroled in the four studies as a representative sample of people seen in clinical practice, 837 of them (31.9%) would have an indeterminate test and require additional diagnostic testing.

Severe fibrosis (≥ F3)

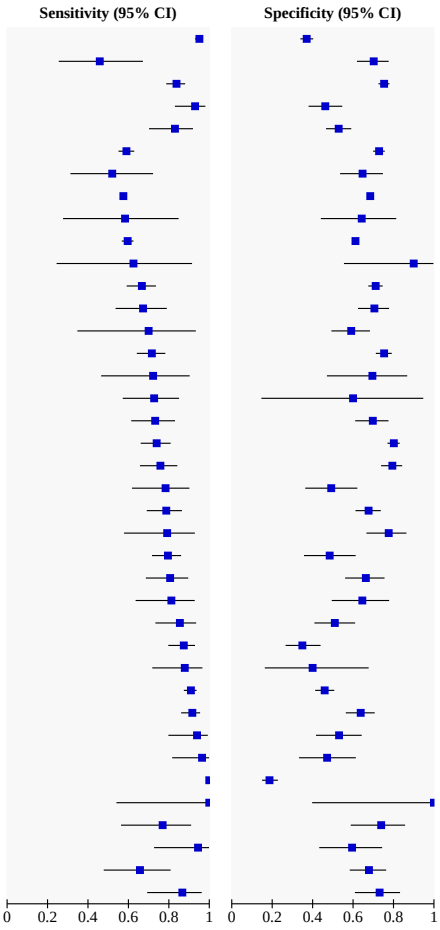
Low cut-off ~ 1.45

Thirty-nine study cohorts (from 29 studies) with 86,907 participants provided data assessing the low cut-off for the FIB-4 score for the diagnosis of severe fibrosis or worse (≥ F3) (Abdel-Hameed 2021a; Abdel-Hameed 2021b; Abdelsameea 2020; Ahmad 2011; Alboraie 2015; Baldwin 2020; Demma 2018; De Oliveira 2016; Ferenci 2014; Fontaine 2009; Fouad 2018; Gamil 2017; Gökan 2016; Hashem 2021; Hseih 2012; Loko 2008; Maheshwari 2013; Martinez 2011; Portilla 2009; Shiha 2022a; Shiha 2022b; Shiha 2022c; Shiha 2022d; Shiha 2022e; Shiha 2022f; Shiha 2022g; Shiha 2022h; Shiha 2022i; Shiha 2022j; Sterling 2006; Stibbe 2011; Tanwar 2017; Trang 2008; Trifan 2009; Tural 2009; Udompap 2020; Vallet-Pichard 2007; Wang 2017; Yilmaz 2021).

The cut-off values used on these study cohorts ranged from 1.3 to 1.60. The validated cut-off of 1.45 was used on 31 occasions (80%). The sensitivity of the low cut-off for the FIB-4 score for the diagnosis of F3 fibrosis or worse ranged from 45.8% to 100%, and the specificity ranged from 19% to 100% (Figure 9).

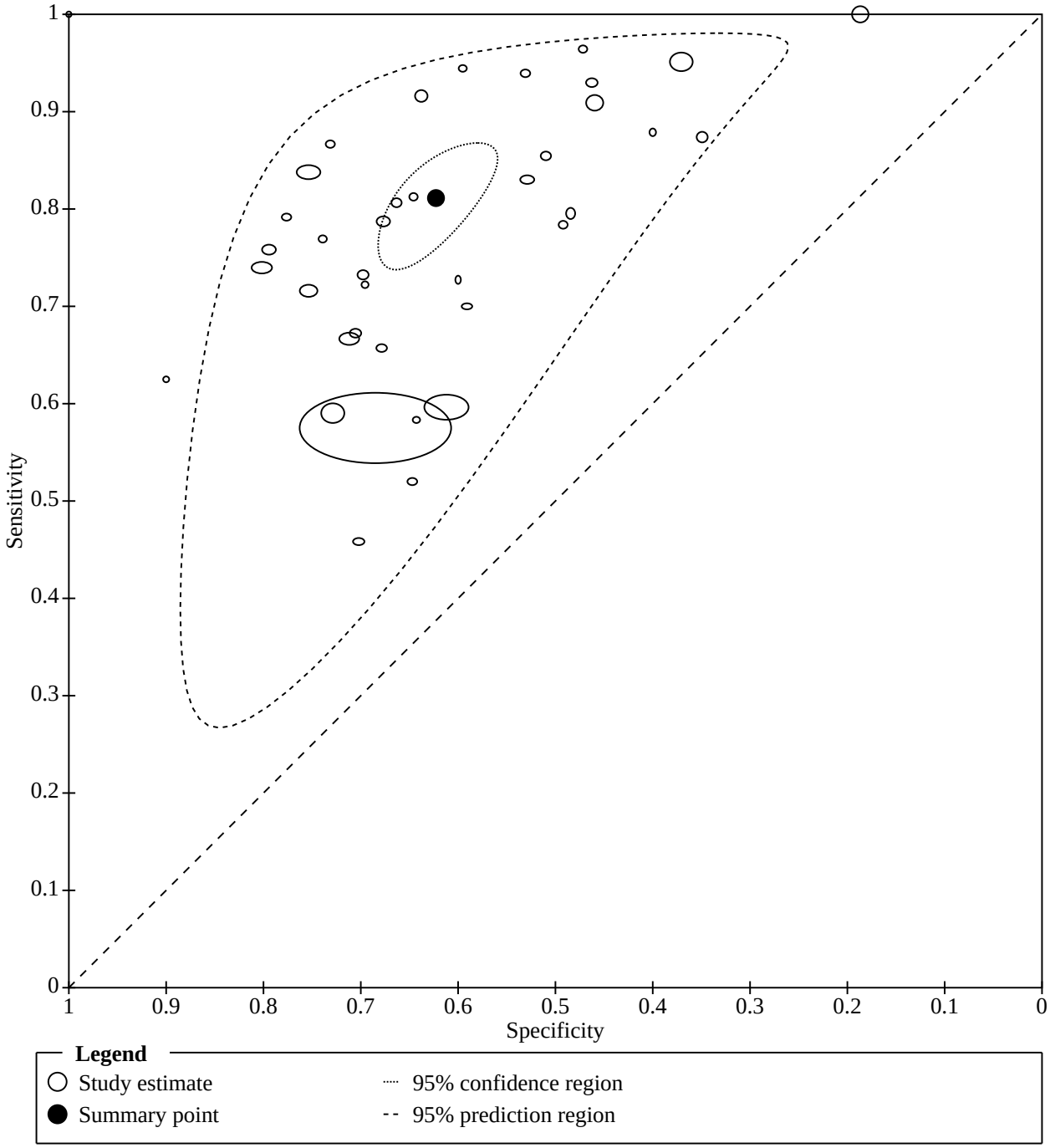
Figure 9. Forest plot of FIB4 for F3 - Studies with cut-off around 1.45.

Study	TP	FP	FN	TN	Cutoff	Sensitivity (95% CI)	Specificity (95% CI)
Wang 2015	526	579	27	341	1	0.95 [0.93, 0.97]	0.37 [0.34, 0.40]
Portilla 2009	11	42	13	99	1.3	0.46 [0.26, 0.67]	0.70 [0.62, 0.78]
Abdelsameea 2020	217	249	42	762	1.3	0.84 [0.79, 0.88]	0.75 [0.73, 0.78]
Maheshwari 2013	53	79	4	68	1.3	0.93 [0.83, 0.98]	0.46 [0.38, 0.55]
Trifan 2009	44	122	9	137	1.36	0.83 [0.70, 0.92]	0.53 [0.47, 0.59]
Ferenci 2014	376	259	261	696	1.4	0.59 [0.55, 0.63]	0.73 [0.70, 0.76]
Fontaine 2009	13	30	12	55	1.45	0.52 [0.31, 0.72]	0.65 [0.54, 0.75]
Hashem 2021	6679	18110	4936	39382	1.45	0.58 [0.57, 0.58]	0.68 [0.68, 0.69]
Shiha 2022h	7	10	5	18	1.45	0.58 [0.28, 0.85]	0.64 [0.44, 0.81]
Shiha 2022a	690	1653	467	2607	1.45	0.60 [0.57, 0.62]	0.61 [0.60, 0.63]
Shiha 2022i	5	1	3	9	1.45	0.63 [0.24, 0.91]	0.90 [0.55, 1.00]
Sterling 2006	116	189	58	467	1.45	0.67 [0.59, 0.74]	0.71 [0.68, 0.75]
Shiha 2022e	39	43	19	103	1.45	0.67 [0.54, 0.79]	0.71 [0.62, 0.78]
Gökan 2016	7	45	3	65	1.45	0.70 [0.35, 0.93]	0.59 [0.49, 0.68]
Gamil 2017	121	119	48	364	1.45	0.72 [0.64, 0.78]	0.75 [0.71, 0.79]
Stibbe 2011	13	7	5	16	1.45	0.72 [0.47, 0.90]	0.70 [0.47, 0.87]
Shiha 2022g	32	2	12	3	1.45	0.73 [0.57, 0.85]	0.60 [0.15, 0.95]
Loko 2008	52	39	19	90	1.45	0.73 [0.61, 0.83]	0.70 [0.61, 0.78]
Vallet-Pichard 2007	108	139	38	562	1.45	0.74 [0.66, 0.81]	0.80 [0.77, 0.83]
Shiha 2022b	69	51	22	197	1.45	0.76 [0.66, 0.84]	0.79 [0.74, 0.84]
Demma 2018	29	32	8	31	1.45	0.78 [0.62, 0.90]	0.49 [0.36, 0.62]
Tural 2009	74	74	20	155	1.45	0.79 [0.69, 0.86]	0.68 [0.61, 0.74]
Alboraie 2015	19	17	5	59	1.45	0.79 [0.58, 0.93]	0.78 [0.67, 0.86]
Shiha 2022f	105	33	27	31	1.45	0.80 [0.72, 0.86]	0.48 [0.36, 0.61]
Shiha 2022c	50	33	12	65	1.45	0.81 [0.69, 0.90]	0.66 [0.56, 0.76]
Tanwar 2017	26	17	6	31	1.45	0.81 [0.64, 0.93]	0.65 [0.49, 0.78]
Ahmad 2011	47	50	8	52	1.45	0.85 [0.73, 0.94]	0.51 [0.41, 0.61]
Hseih 2012	97	82	14	44	1.45	0.87 [0.80, 0.93]	0.35 [0.27, 0.44]
Baldwin 2020	29	9	4	6	1.45	0.88 [0.72, 0.97]	0.40 [0.16, 0.68]
De Oliveira 2016	320	241	32	205	1.45	0.91 [0.87, 0.94]	0.46 [0.41, 0.51]
Martinez 2011	142	67	13	118	1.45	0.92 [0.86, 0.95]	0.64 [0.56, 0.71]
Yilmaz 2021	31	38	2	43	1.45	0.94 [0.80, 0.99]	0.53 [0.42, 0.64]
Trang 2008	27	28	1	25	1.45	0.96 [0.82, 1.00]	0.47 [0.33, 0.61]
Shiha 2022d	390	331	0	76	1.45	1.00 [0.99, 1.00]	0.19 [0.15, 0.23]
Shiha 2022j	6	0	0	4	1.45	1.00 [0.54, 1.00]	1.00 [0.40, 1.00]
Fouad 2018	20	12	6	34	1.5	0.77 [0.56, 0.91]	0.74 [0.59, 0.86]
Udompap 2020	17	17	1	25	1.57	0.94 [0.73, 1.00]	0.60 [0.43, 0.74]
Abdel-Hameed 2021b	23	36	12	76	2	0.66 [0.48, 0.81]	0.68 [0.58, 0.76]
Abdel-Hameed 2021a	26	18	4	49	2	0.87 [0.69, 0.96]	0.73 [0.61, 0.83]



We combined these data in a meta-analysis, and obtained the following summary estimates for the FIB-4 score using a low cut-off (~ 1.45) for the diagnosis of severe fibrosis (≥ F3): sensitivity 81.1% (95% CI 75.6% to 85.6%); specificity 62.3% (95% CI 57.4% to 66.9%); LR+ 2.2 (95% CI 2.0 to 2.4); LR- 0.30 (95% CI 0.24 to 0.38) (Figure 10).

Figure 10. Summary ROC plot of FIB4 for F3 - Studies with cut-off around 1.45. The circles represent individual studies.
The solid circle represents the summary estimate of sensitivity and specificity.
The dotted line represents the 95% confidence regions.
The dashed line represents the 95% prediction regions



Amongst these study cohorts, the median prevalence of severe fibrosis was 30.9% (IQR 23.8% to 43.9%). Using this value as a pre-test probability, we obtained a post-test probability of 49.6% (95% CI 47.2% to 51.8%) when the test was positive and a post-test probability of 11.8% (95% CI 9.7% to 14.5%) when the test was negative.

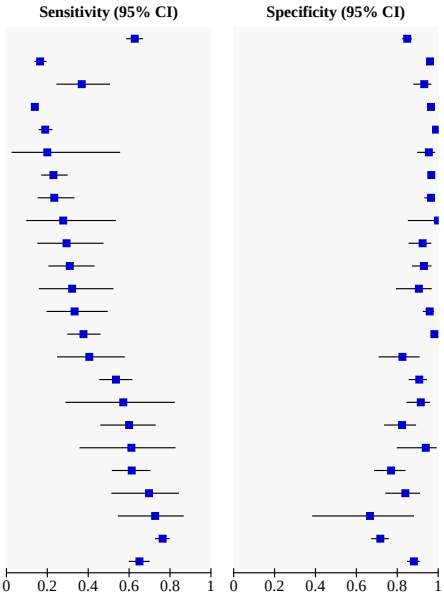
High cut-off ~ 3.25

Twenty-four study cohorts (from 24 studies) with 81,350 participants provided data assessing the high cut-off for the FIB-4 score for the diagnosis of severe fibrosis or worse (≥ F3) (Ahmad 2011; Attallah 2012; Baldwin 2020; Demma 2018; De Oliveira 2016; Ferenci 2014; Gökan 2016; Gounder 2018; Hashem 2021; Hseih 2012; Ikatura 2021; Loko 2008; Maheshwari 2013; Martinez 2011; Qian 2019; Schmoyer 2020; Silva Junior 2014; Sterling 2006; Stibbe 2011; Trang 2008; Tural 2009; Vallet-Pichard 2007; Wang 2015; Yilmaz 2021).

The cut-off values used on these study cohorts ranged from 3.20 to 3.27. The validated cut-off of 3.25 was used on 20 occasions (83%). The sensitivity of the high cut-off for the FIB-4 score for the diagnosis of F3 fibrosis or worse ranged from 13.9% to 76.5%, and the specificity ranged from 66.7% to 100% (Figure 11).

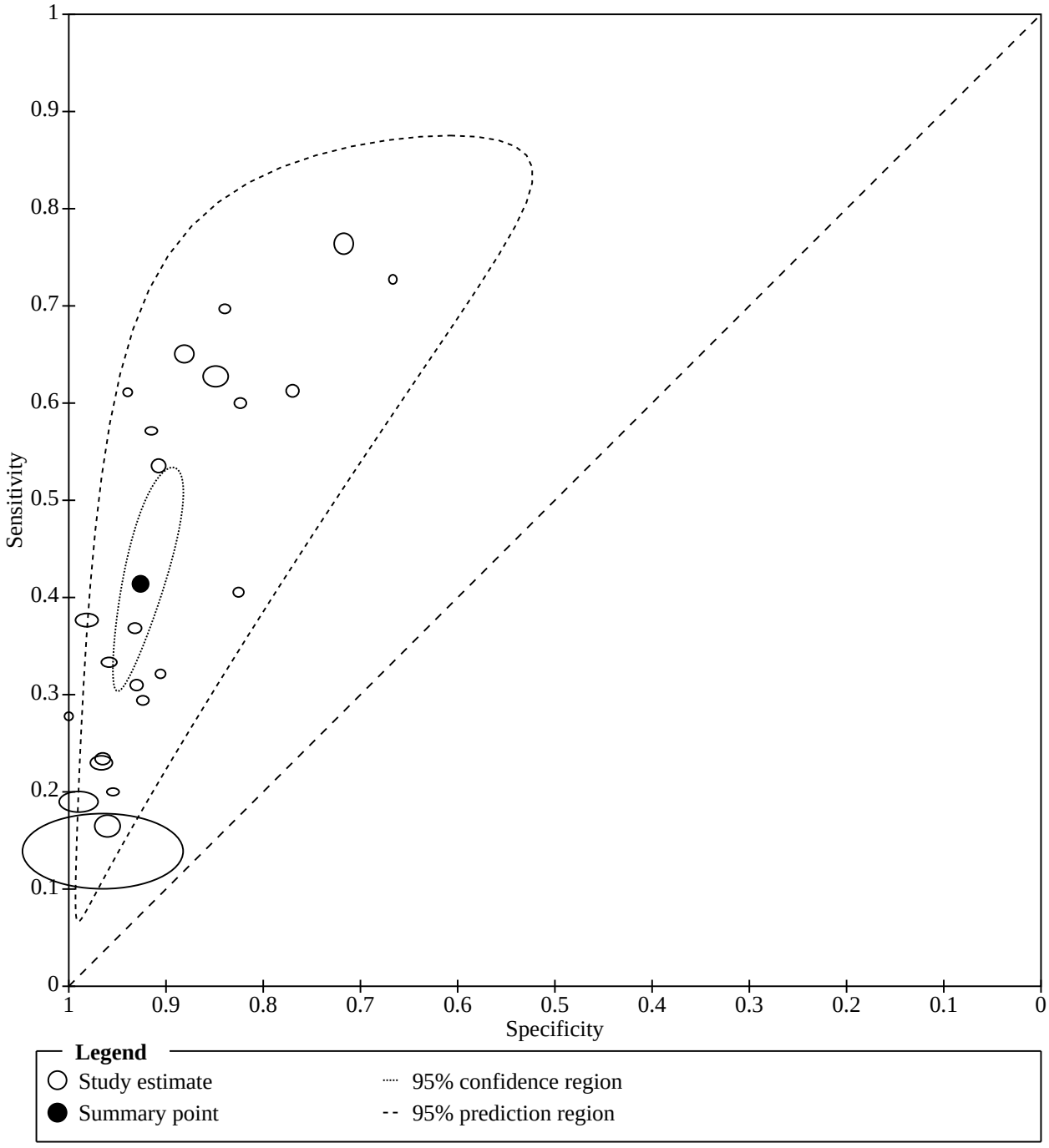
Figure 11. Forest plot of FIB4 for F3 - Studies with cut-off around 3.25.

Study	TP	FP	FN	TN	Cutoff	Sensitivity (95% CI)	Specificity (95% CI)
Wang 2015	347	139	206	781	3	0.63 [0.59, 0.67]	0.85 [0.82, 0.87]
Ferenci 2014	105	38	532	917	3.2	0.16 [0.14, 0.20]	0.96 [0.95, 0.97]
Maheshwari 2013	21	10	36	137	3.23	0.37 [0.24, 0.51]	0.93 [0.88, 0.97]
Hashem 2021	1614	2012	10001	55480	3.25	0.14 [0.13, 0.15]	0.97 [0.96, 0.97]
Attallah 2012	104	27	444	2637	3.25	0.19 [0.16, 0.23]	0.99 [0.99, 0.99]
Gökan 2016	2	5	8	105	3.25	0.20 [0.03, 0.56]	0.95 [0.90, 0.99]
Sterling 2006	40	22	134	634	3.25	0.23 [0.17, 0.30]	0.97 [0.95, 0.98]
Tural 2009	22	8	72	221	3.25	0.23 [0.15, 0.33]	0.97 [0.93, 0.98]
Stibbe 2011	5	0	13	23	3.25	0.28 [0.10, 0.53]	1.00 [0.85, 1.00]
Schmoyer 2020	10	8	24	97	3.25	0.29 [0.15, 0.47]	0.92 [0.86, 0.97]
Loko 2008	22	9	49	120	3.25	0.31 [0.21, 0.43]	0.93 [0.87, 0.97]
Trang 2008	9	5	19	48	3.25	0.32 [0.16, 0.52]	0.91 [0.79, 0.97]
Gounder 2018	14	10	28	231	3.25	0.33 [0.20, 0.50]	0.96 [0.93, 0.98]
Vallet-Pichard 2007	55	13	91	688	3.25	0.38 [0.30, 0.46]	0.98 [0.97, 0.99]
Demma 2018	15	11	22	52	3.25	0.41 [0.25, 0.58]	0.83 [0.71, 0.91]
Martinez 2011	83	17	72	167	3.25	0.54 [0.45, 0.62]	0.91 [0.86, 0.95]
Qian 2019	8	9	6	97	3.25	0.57 [0.29, 0.82]	0.92 [0.84, 0.96]
Ahmad 2011	33	18	22	84	3.25	0.60 [0.46, 0.73]	0.82 [0.74, 0.89]
Silva Junior 2014	11	2	7	31	3.25	0.61 [0.36, 0.83]	0.94 [0.80, 0.99]
Hseih 2012	68	29	43	97	3.25	0.61 [0.52, 0.70]	0.77 [0.69, 0.84]
Yilmaz 2021	23	13	10	68	3.25	0.70 [0.51, 0.84]	0.84 [0.74, 0.91]
Baldwin 2020	24	5	9	10	3.25	0.73 [0.54, 0.87]	0.67 [0.38, 0.88]
Ikatura 2021	437	125	135	317	3.2565	0.76 [0.73, 0.80]	0.72 [0.67, 0.76]
De Oliveira 2016	229	53	123	393	3.27	0.65 [0.60, 0.70]	0.88 [0.85, 0.91]



We combined these data in a meta-analysis, and obtained the following summary estimates for the FIB-4 score using a high cut-off (~ 3.25) for the diagnosis of severe fibrosis (≥ F3): sensitivity 41.4% (95% CI 33.0% to 50.4%); specificity 92.6% (95% CI 89.5% to 94.9%); LR+ 5.6 (95% CI 4.4 to 7.1); LR- 0.63 (95% CI 0.56 to 0.72) (Figure 12).

Figure 12. Summary ROC plot of FIB4 for F3 - Studies with cut-off around 3.25. The circles represent individual studies.
The solid circle represents the summary estimate of sensitivity and specificity.
The dotted line represents the 95% confidence regions.
The dashed line represents the 95% prediction regions



Amongst these study cohorts, the median prevalence of severe fibrosis was 34.8% (IQR 19.1% to 42.5%). Using this value as a pre-test probability, we obtained a post-test probability of 74.9% (95% CI 70.1% to 79.1%) when the test was positive and a post-test probability of 25.2% (95% CI 23.0% to 27.8%) when the test was negative.

Indeterminate results (grey area)

This represents the number of people for whom the test gives an inconclusive result (i.e. cannot rule in using a high cut-off and cannot rule out using a low cut-off). When calculating this grey area, we considered only the 18 studies reporting both the high and low cut-offs when applying the FIB-4 score to severe fibrosis (Ahmad 2011; Baldwin 2020; Demma 2018; De Oliveira 2016; Ferenci 2014; Gökan 2016; Hashem 2021; Hseih 2012; Loko 2008; Maheshwari 2013; Martinez 2011; Sterling 2006; Stibbe 2011; Trang 2008; Tural 2009; Vallet-Pichard 2007; Wang 2015; Yilmaz 2021). If we consider all 76,612 participants enroled in these 18 studies as a representative sample of people seen in clinical practice, 23,669 of them (30.9%) would have an indeterminate test and require additional diagnostic testing.

Cirrhosis (F4)

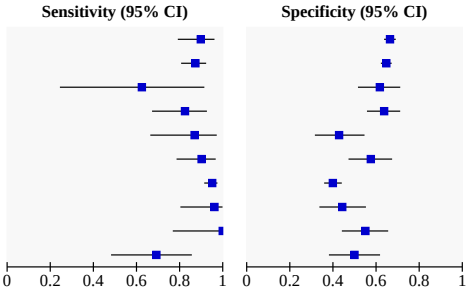
Low cut-off ~ 1.45

Ten study cohorts (from 10 studies) with 4537 participants provided data assessing the low cut-off for the FIB-4 score for the diagnosis of cirrhosis (F4) (Abdelsameea 2020; Abo El-Khair 2019; Demma 2018; De Oliveira 2016; Ferenci 2014; Fontaine 2009; Güzelbulut 2011; Loko 2008; Schmid 2015; Yilmaz 2021).

The cut-off values used on these study cohorts ranged from 1.30 to 1.60. The validated cut-off of 1.45 was used on seven occasions (70%). The sensitivity of the low cut-off for the FIB-4 score for the diagnosis of cirrhosis (F4) ranged from 57% to 100%, and the specificity ranged from 40% to 66% (Figure 13).

Figure 13. Forest plot of FIB4 for F4 - Studies with cut-off around 1.45.

Study	TP	FP	FN	TN	Cutoff	Sensitivity (95% CI)	Specificity (95% CI)
Abdelsameea 2020	53	406	6	805	1.3	0.90 [0.79, 0.96]	0.66 [0.64, 0.69]
Ferenci 2014	124	511	18	939	1.4	0.87 [0.81, 0.92]	0.65 [0.62, 0.67]
Fontaine 2009	5	39	3	63	1.45	0.63 [0.24, 0.91]	0.62 [0.52, 0.71]
Loko 2008	33	58	7	102	1.45	0.82 [0.67, 0.93]	0.64 [0.56, 0.71]
Demma 2018	20	44	3	33	1.45	0.87 [0.66, 0.97]	0.43 [0.32, 0.55]
Güzelbulut 2011	46	42	5	57	1.45	0.90 [0.79, 0.97]	0.58 [0.47, 0.67]
De Oliveira 2016	214	344	11	229	1.45	0.95 [0.91, 0.98]	0.40 [0.36, 0.44]
Yilmaz 2021	25	49	1	39	1.45	0.96 [0.80, 1.00]	0.44 [0.34, 0.55]
Schmid 2015	14	40	0	49	1.45	1.00 [0.77, 1.00]	0.55 [0.44, 0.66]
Abo El-Khair 2019	18	37	8	37	2	0.69 [0.48, 0.86]	0.50 [0.38, 0.62]



We combined these data in a meta-analysis, and obtained the following summary estimates for the FIB-4 score using a low cut-off (~ 1.45) for the diagnosis of cirrhosis (F4): sensitivity 89.1% (95% CI 83.9% to 92.8%); specificity 55.6% (95% CI 49.3% to 61.7%); LR+ 2.0 (95% CI 1.8 to 2.3); LR- 0.20 (95% CI 0.14 to 0.28).

Amongst these study cohorts, the median prevalence of cirrhosis was 21.4% (IQR 8.9% to 26.0%). Using this value as a pre-test probability, we obtained a post-test probability of 35.3% (95% CI 32.9% to 38.5%) when the test was positive and a post-test probability of 5.2% (95% CI 3.7% to 7.1%) when the test was negative.

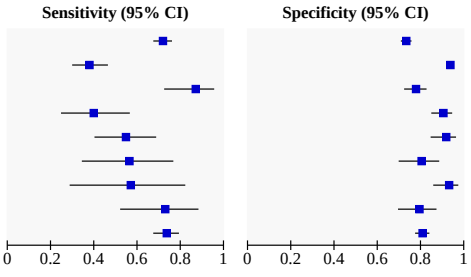
High cut-off ~ 3.25

Nine study cohorts (from nine studies) with 5075 participants provided data assessing the high cut-off for the FIB-4 score for the diagnosis of cirrhosis (F4) (Demma 2018; De Oliveira 2016; Ferenci 2014; Güzelbulut 2011; Loko 2008; Schmid 2015; Tsukano 2017; Yen 2018; Yilmaz 2021).

The cut-off values used on these study cohorts ranged from 3.1 to 3.27. The validated cut-off of 3.25 was used on five occasions (56%). The sensitivity of the high cut-off for the FIB-4 score for the diagnosis of cirrhosis (F4) ranged from 38% to 87%, and the specificity ranged from 73% to 94% (Figure 14).

Figure 14. Forest plot of FIB4 for F4 - Studies with cut-off around 3.25.

Study	TP	FP	FN	TN	Cutoff	Sensitivity (95% CI)	Specificity (95% CI)
Yen 2018	314	340	122	940	3.1	0.72 [0.68, 0.76]	0.73 [0.71, 0.76]
Ferenci 2014	54	89	88	1361	3.2	0.38 [0.30, 0.47]	0.94 [0.93, 0.95]
Tsukano 2017	34	58	5	205	3.24	0.87 [0.73, 0.96]	0.78 [0.72, 0.83]
Loko 2008	16	15	24	145	3.25	0.40 [0.25, 0.57]	0.91 [0.85, 0.95]
Güzelbulut 2011	28	8	23	91	3.25	0.55 [0.40, 0.69]	0.92 [0.85, 0.96]
Demma 2018	13	15	10	62	3.25	0.57 [0.34, 0.77]	0.81 [0.70, 0.89]
Schmid 2015	8	6	6	83	3.25	0.57 [0.29, 0.82]	0.93 [0.86, 0.97]
Yilmaz 2021	19	18	7	70	3.25	0.73 [0.52, 0.88]	0.80 [0.70, 0.87]
De Oliveira 2016	166	109	59	464	3.27	0.74 [0.68, 0.79]	0.81 [0.78, 0.84]



We combined these data in a meta-analysis, and obtained the following summary estimates for the FIB-4 score using a high cut-off (~ 3.25) for the diagnosis of cirrhosis (F4): sensitivity 61.2% (95% CI 50.7% to 70.8%); specificity 85.9% (95% CI 80.2% to 90.2%); LR+ 4.4 (95% CI 3.4 to 5.5); LR- 0.45 (95% CI 0.36 to 0.56).

Amongst these study cohorts, the median prevalence of cirrhosis was 22.8% (IQR 13.6% to 25.4%). Using this value as a pre-test probability, we obtained a post-test probability of 56.5% (95% CI 50.1% to 61.9%) when the test was positive and a post-test probability of 11.7% (95% CI 9.6% to 4.2%) when the test was negative.

Indeterminate results (grey area)

This represents the number of people for whom the test gives an inconclusive result (i.e. cannot rule in using a high cut-off and cannot rule out using a low cut-off). To calculate this grey area, we considered only the seven studies reporting both the high and low cut-offs when applying the FIB-4 score to cirrhosis (Demma 2018; De Oliveira 2016; Ferenci 2014; Güzelbulut 2011; Loko 2008; Schmid 2015; Yilmaz 2021). If we consider all 3057 people in these seven studies as a representative sample of people seen in clinical practice, we find that 1000 of them (32.7%) would have had an indeterminate result when using a dual cut-off and would therefore require additional diagnostic testing.

Forns index

Significant fibrosis (≥ F2)

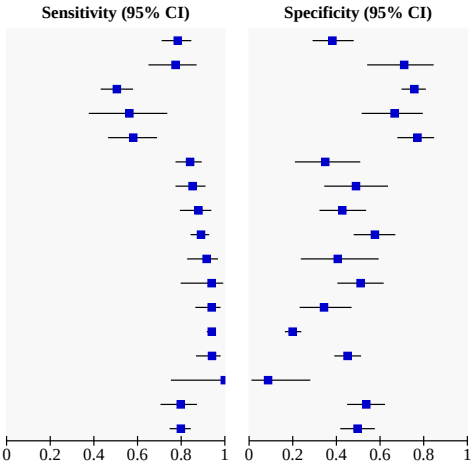
Low cut-off ~ 4.2

Seventeen study cohorts (from 15 studies) with 4354 participants provided data assessing the low cut-off for the Forns index for the diagnosis of significant fibrosis (≥ F2) (Alboraie 2015; Bourliere 2006; Corradi 2009; Crisan 2012; Forns 2002a; Forns 2002b; Güzelbulut 2011; Koller 2014; Ladero 2010; Leroy 2007; Loko 2008; Macías 2006; Martinez 2011; Mossong 2011; Sebastiani 2008a; Sebastiani 2008b; Sebastiani 2012).

The cut-off values used on these study cohorts ranged from 4.02 to 4.47. The validated cut-off of 4.2 was used on 14 occasions (56%). The sensitivity of the low cut-off for the Forns index for the diagnosis of significant fibrosis (F2) ranged from 54% to 100%, and specificity ranged from 8% to 80% (Figure 15).

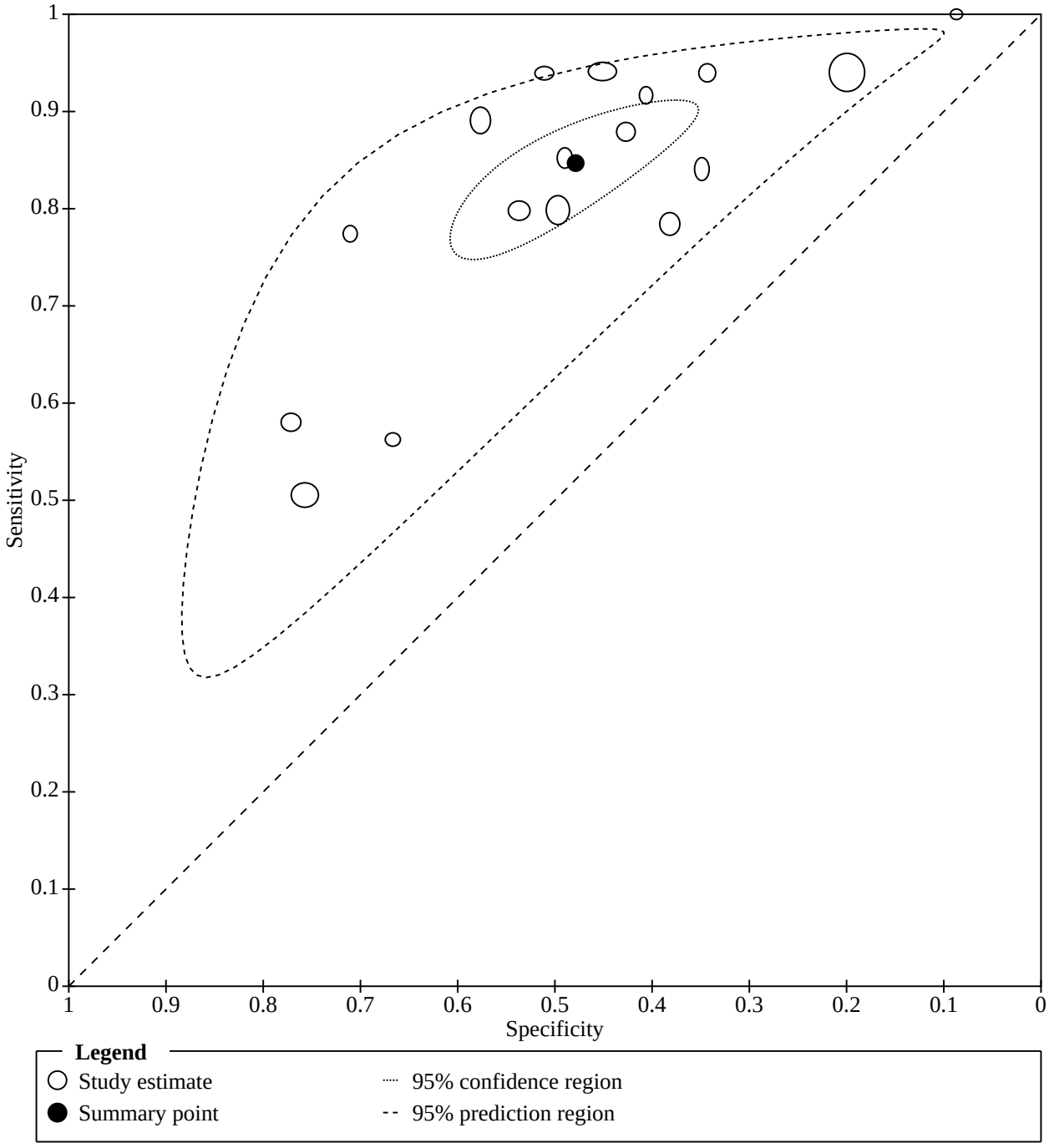
Figure 15. Forest plot of Forns for F2 - Studies with cut-off around 4.2.

Study	TP	FP	FN	TN	Cutoff	Sensitivity (95% CI)	Specificity (95% CI)
Macías 2006	120	68	33	42	4	0.78 [0.71, 0.85]	0.38 [0.29, 0.48]
Alboraie 2015	48	11	14	27	4.02	0.77 [0.65, 0.87]	0.71 [0.54, 0.85]
Ladero 2010	94	59	92	184	4.2	0.51 [0.43, 0.58]	0.76 [0.70, 0.81]
Sebastiani 2008a	18	16	14	32	4.2	0.56 [0.38, 0.74]	0.67 [0.52, 0.80]
Mossong 2011	47	24	34	81	4.2	0.58 [0.47, 0.69]	0.77 [0.68, 0.85]
Loko 2008	132	28	25	15	4.2	0.84 [0.77, 0.89]	0.35 [0.21, 0.51]
Sebastiani 2008b	98	25	17	24	4.2	0.85 [0.77, 0.91]	0.49 [0.34, 0.64]
Leroy 2007	80	51	11	38	4.2	0.88 [0.79, 0.94]	0.43 [0.32, 0.54]
Martinez 2011	204	47	25	64	4.2	0.89 [0.84, 0.93]	0.58 [0.48, 0.67]
Koller 2014	66	19	6	13	4.2	0.92 [0.83, 0.97]	0.41 [0.24, 0.59]
Forns 2002b	31	45	2	47	4.2	0.94 [0.80, 0.99]	0.51 [0.40, 0.62]
Güzelbulut 2011	78	44	5	23	4.2	0.94 [0.86, 0.98]	0.34 [0.23, 0.47]
Sebastiani 2012	519	369	33	92	4.2	0.94 [0.92, 0.96]	0.20 [0.16, 0.24]
Forns 2002a	80	146	5	120	4.2	0.94 [0.87, 0.98]	0.45 [0.39, 0.51]
Corradi 2009	13	21	0	2	4.2	1.00 [0.75, 1.00]	0.09 [0.01, 0.28]
Bourliere 2006	79	63	20	73	4.21	0.80 [0.71, 0.87]	0.54 [0.45, 0.62]
Crisan 2012	226	82	57	81	4.47	0.80 [0.75, 0.84]	0.50 [0.42, 0.58]



We combined these data in a meta-analysis, and obtained the following summary estimates for the use of the Forns index with a low cut-off (~ 4.2) for the diagnosis of significant fibrosis (≥ F2): sensitivity 84.7% (95% CI 77.9% to 89.7%); specificity 47.9% (95% CI 38.6% to 57.3%); LR+ 1.6 (95% CI 1.4 to 1.9); LR- 0.32 (95% CI 0.25 to 0.41) (Figure 16).

Figure 16. Summary ROC plot of Forns for F2 - Studies with cut-off around 4.2. The circles represent individual studies.
The solid circle represents the summary estimate of sensitivity and specificity.
The dotted line represents the 95% confidence regions.
The dashed line represents the 95% prediction regions.



Amongst these study cohorts, the median prevalence of significant fibrosis was 54.5% (IQR 40.0% to 62.0%). Using this value as a pre-test probability, we obtained a post-test probability of 65.7% (95% CI 62.6% to 69.5%) when the test was positive and a post-test probability of 27.7% (95% CI 23.0% to 32.9%) when the test was negative.

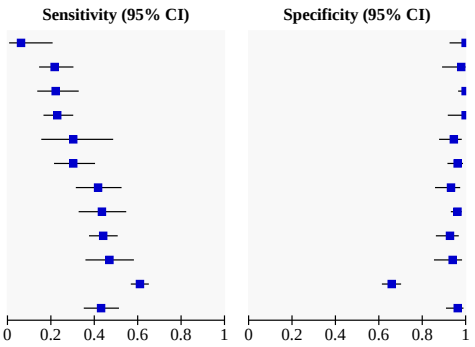
High cut-off ~ 6.9

Twelve study cohorts (from 10 studies) with 3245 participants provided data assessing the high cut-off for the Forns index for the diagnosis of significant fibrosis (≥ F2) (Bourliere 2006; Forns 2002a; Forns 2002b; Güzelbulut 2011; Leroy 2007; Loko 2008; Macías 2006; Martinez 2011; Mossong 2011; Sebastiani 2008a; Sebastiani 2008b; Sebastiani 2012).

All studies used the validated 6.9 cut-off. The sensitivity of the high cut-off for the Forns index for the diagnosis of significant fibrosis (≥ F2) ranged from 5% to 61%, and the specificity ranged from 8% to 80% (Figure 17).

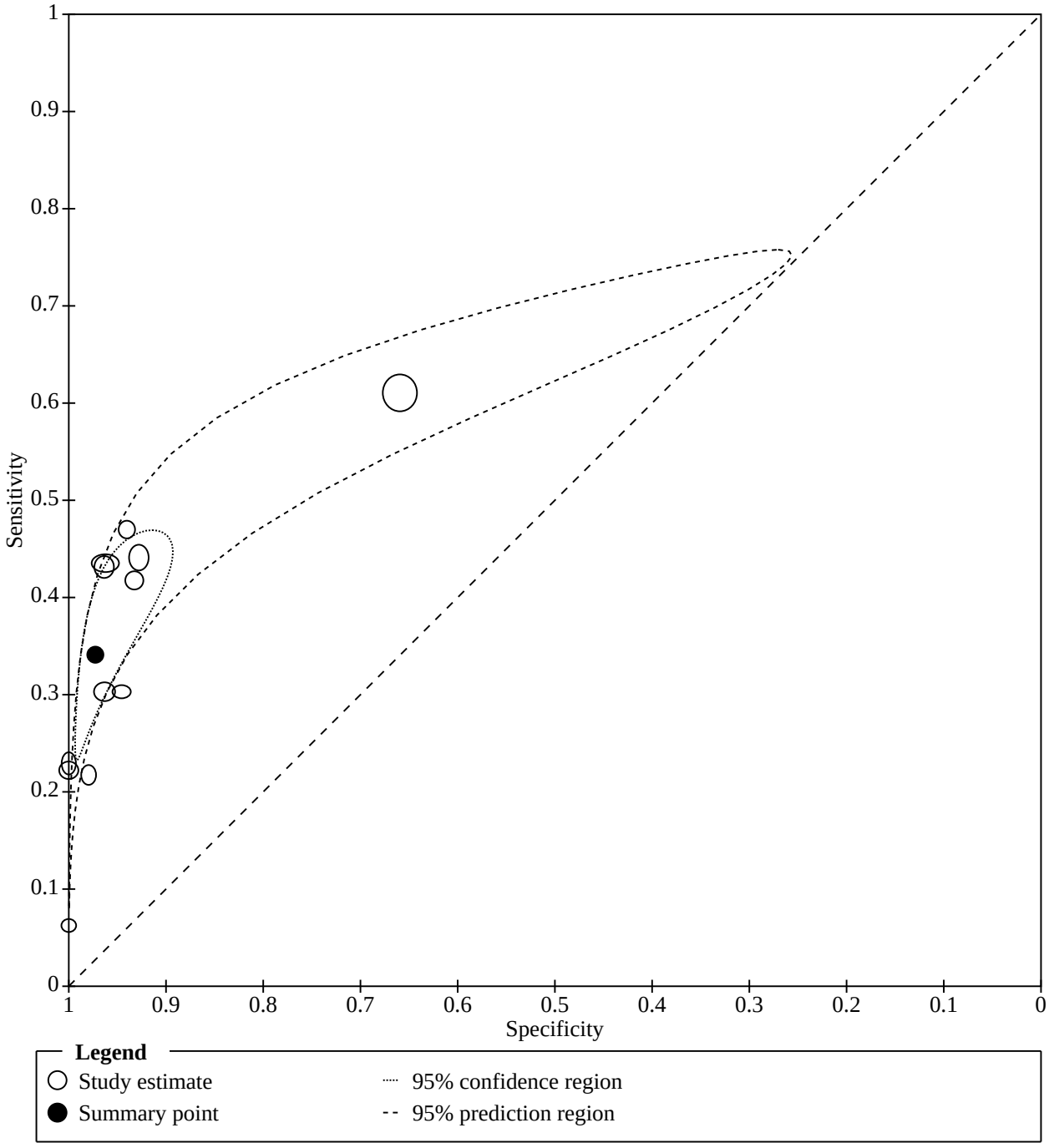
Figure 17. Forest plot of Forns for F2 - Studies with cut-off around 6.9.

Study	TP	FP	FN	TN	Cutoff	Sensitivity (95% CI)	Specificity (95% CI)
Sebastiani 2008a	2	0	30	48	6.9	0.06 [0.01, 0.21]	1.00 [0.93, 1.00]
Sebastiani 2008b	25	1	90	48	6.9	0.22 [0.15, 0.30]	0.98 [0.89, 1.00]
Mossong 2011	18	0	63	105	6.9	0.22 [0.14, 0.33]	1.00 [0.97, 1.00]
Loko 2008	36	0	121	43	6.9	0.23 [0.17, 0.30]	1.00 [0.92, 1.00]
Forns 2002b	10	5	23	87	6.9	0.30 [0.16, 0.49]	0.95 [0.88, 0.98]
Bourliere 2006	30	5	69	131	6.9	0.30 [0.21, 0.40]	0.96 [0.92, 0.99]
Leroy 2007	38	6	53	83	6.9	0.42 [0.32, 0.53]	0.93 [0.86, 0.97]
Forns 2002a	37	10	48	256	6.9	0.44 [0.33, 0.55]	0.96 [0.93, 0.98]
Martinez 2011	101	8	128	103	6.9	0.44 [0.38, 0.51]	0.93 [0.86, 0.97]
Güzelbulut 2011	39	4	44	63	6.9	0.47 [0.36, 0.58]	0.94 [0.85, 0.98]
Sebastiani 2012	337	157	215	304	6.9	0.61 [0.57, 0.65]	0.66 [0.61, 0.70]
Macías 2006	66	4	87	106	7	0.43 [0.35, 0.51]	0.96 [0.91, 0.99]



We combined these data in a meta-analysis, and obtained the following summary estimates for the use of the Forns index with a high cut-off (~ 6.9) for diagnosing significant fibrosis (≥ F2): sensitivity 34.1% (95% CI 26.4% to 42.8%); specificity 97.3% (95% CI 92.9% to 99.0%); LR+ 12.5 (95% CI 5.7 to 27.2); LR- 0.68 (95% CI 0.61 to 0.75) (Figure 18).

Figure 18. Summary ROC plot of Forns for F2 - Studies with cut-off around 6.9. The circles represent individual studies.
The solid circle represents the summary estimate of sensitivity and specificity.
The dotted line represents the 95% confidence regions.
The dashed line represents the 95% prediction regions.



Amongst these study cohorts, the median prevalence of significant fibrosis was 52.5% (IQR 34.2% to 58.2%). Using this value as a pre-test probability, we obtained a post-test probability of 93.3% (95% CI 86.3% to 96.8%) when the test was positive and a post-test probability of 42.9% (95% CI 40.3% to 45.3%) when the test was negative.

Indeterminate result (grey area)

This represents the number of people for whom the test gives an inconclusive result (i.e. cannot rule in using a high cut-off and cannot rule out using a low cut-off). To calculate this grey area, we considered only the 12 studies reporting both the high and low cut-offs when applying the Forns index to significant fibrosis (Bourliere 2006; Forns 2002a; Forns 2002b; Güzelbulut 2011; Leroy 2007; Loko 2008; Macías 2006; Martinez 2011; Mossong 2011; Sebastiani 2008a; Sebastiani 2008b; Sebastiani 2012). If we consider all 3245 participants in these 12 studies as a representative sample of people seen in clinical practice, we find that 1473 of them (44.8%) would have had an indeterminate result when using a dual cut-off, and would therefore require additional diagnostic testing.

Severe fibrosis (≥ F3)

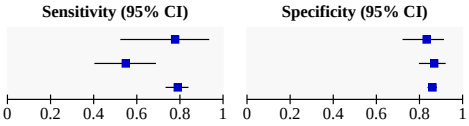
High cut-off ~ 6.9

Three study cohorts (from three studies) with 1516 participants provided data assessing the high cut-off for the Forns index for diagnosing severe fibrosis (≥ F3) (Andrés-Otero 2016; Iacobellis 2005; Leroy 2007). We could not use the bivariate model to perform a meta-analysis as fewer than four studies were included.

The cut-off values used on these study cohorts were 6.66, 6.9, and 6.9, respectively. The sensitivity of the high cut-off for the Forns index for diagnosing severe fibrosis (≥ F3) ranged from 55% to 79%, and the specificity ranged from 83% to 87% (Figure 19).

Figure 19. Forest plot of Forns for F3 - Studies with cut-off around 6.9.

Study	TP	FP	FN	TN	Cutoff	Sensitivity (95% CI)	Specificity (95% CI)
Andrés-Otero 2016	14	11	4	55	6.66	0.78 [0.52, 0.94]	0.83 [0.72, 0.91]
Leroy 2007	28	17	23	112	6.9	0.55 [0.40, 0.69]	0.87 [0.80, 0.92]
Iacobellis 2005	192	141	51	868	6.9	0.79 [0.73, 0.84]	0.86 [0.84, 0.88]



Indeterminate result (grey area)

This represents the number of people for whom the test gives an inconclusive result (i.e. cannot rule in using a high cut-off and cannot rule out using a low cut-off). To calculate this grey area, we considered only one study reporting both the high and low cut-offs when applying the Forns index to severe fibrosis (Leroy 2007). In this single study, 87 of 180 people (48.3%) would have had an indeterminate result when using a dual cut-off and would therefore require additional diagnostic testing (Leroy 2007).

Cirrhosis (F4)

High cut-off ~ 6.9

Two study cohorts (from two studies) with 234 participants provided data assessing the high cut-off for the Forns index for diagnosing cirrhosis (F4) (Andrés-Otero 2016; Güzelbulut 2011). We could not use the bivariate model to perform a meta-analysis as fewer than four studies were included.

The cut-off values used on these study cohorts were 6.66 and 6.9, respectively. The sensitivity of the high cut-off for the Forns index for diagnosing cirrhosis (F4) ranged from 67% to 86%, and the specificity ranged from 81% to 91%.

Indeterminate result (grey area)

This represents the number of people for whom the test gives an inconclusive result (i.e. cannot rule in using a high cut-off and cannot rule out using a low cut-off). To calculate this grey area, we considered only one study reporting both the high and low cut-offs when applying the Forns index to diagnosing cirrhosis (Güzelbulut 2011). In this study, 79 (52.7%) of 150 people would have had an indeterminate result when using a dual cut-off and would therefore require additional diagnostic testing (Güzelbulut 2011).

Comparative analysis of the FIB-4 score versus the Forns index

Due to the paucity of data, especially for the Forns index in the diagnosis of severe fibrosis or worse (≥ F3) and cirrhosis (F4), we could only compare the FIB-4 score to the Forns index for diagnosing significant fibrosis (≥ F2).

Low cut-off FIB-4 versus low cut-off Forns index for diagnosing significant fibrosis (≥ F2)

Using a bivariate model, we made an indirect comparison between the studies reporting the diagnostic accuracy of the FIB-4 score and the Forns index at their respective low cut-offs for diagnosing significant fibrosis (≥ F2). This involved 17 studies (5098 participants) and 17 studies (4402 participants) for the FIB-4 score and Forns index, respectively. Overall, we found a statistically significant difference in accuracy between the two tests (P = 0.004). In particular, the Forns index has a slightly higher sensitivity (relative sensitivity 1.12, 95% CI 1.00 to 1.25; P = 0.0573), and lower specificity (relative specificity 0.69, 95% CI 0.57 to 0.84; P = 0.002; Table 3) than the FIB-4 score.

Using a bivariate model, we made an indirect comparison between the studies reporting the diagnostic accuracy of the FIB-4 score and the Forns index at their respective high cut-offs for diagnosing significant fibrosis (≥ F2). This involved five studies (2673 participants) and 12 studies (3287 participants) for the FIB-4 score and Forns index, respectively. Overall, we found no differences in accuracy between the two tests (P = 0.975).

Sensitivity analysis

We obtained pooled estimates restricted to studies at low risk of bias for each QUADAS-2 domain, and only studies published as full texts, and report the results in Table 4. We also investigated changes in the diagnostic accuracy of both tests in every target condition by excluding studies at high or unclear risk of bias. The results in these sensitivity analyses were similar to the main ones.

Heterogeneity analysis

Of our four prespecified heterogeneity analyses (see Investigations of heterogeneity), we had sufficient data to perform three secondary analyses – mean ALT levels; presence or absence of people living with HIV; liver biopsy quality – for two test/target condition/cut-off combinations. We summarise these data in Table 5.

Studies with mean ALT values above 80 IU/L versus 80 IU/L or lower

An ALT level above 80 IU/L represents two times the upper limit of normal.

We performed a heterogeneity analysis for both cut-offs of the FIB-4 score for diagnosing significant fibrosis (≥ F2), severe fibrosis (≥ F3), and cirrhosis (F4). The model did not converge for any of the cut-offs for the diagnosis of significant fibrosis. However, the sensitivity and specificity of the FIB-4 score for the diagnosis of severe fibrosis and cirrhosis were similar when comparing studies with mean ALT values above 80 IU/L to those with mean ALT values of 80 IU/L or lower, indicating that this is likely not a source of heterogeneity.

We performed a heterogeneity analysis for both cut-offs of the Forns index for diagnosing significant fibrosis (≥ F2). The sensitivity and specificity of the Forns index were similar when comparing studies with mean ALT values above 80 IU/L to those with mean ALT values of 80 IU/L or lower, indicating that this is likely not a source of heterogeneity.

Studies with and without people living with HIV

Ten studies included one or more people living with HIV (Abdel-Hameed 2021b; Calès 2010; Guilabert 2010; Loko 2008; Macías 2006; Portilla 2009; Schmid 2015; Sterling 2006; Trang 2008; Tural 2009).

We performed a heterogeneity analysis for both cut-offs of the FIB-4 score for diagnosing significant fibrosis (≥ F2), severe fibrosis (≥ F3), and cirrhosis (F4). The sensitivity and specificity of the FIB-4 score for the diagnosis of any stage of fibrosis were similar when comparing studies with and without people living with HIV, indicating that this is likely not a source of heterogeneity.

We performed a heterogeneity analysis for both cut-offs of the Forns index for diagnosing significant fibrosis (≥ F2). The sensitivity and specificity of the Forns index for diagnosing significant fibrosis were similar when comparing studies with and without people living with HIV, indicating this is likely not a source of heterogeneity.

Studies with high-quality versus low-quality liver biopsy

We planned to perform a heterogeneity analysis comparing studies with high-quality liver biopsy (i.e. biopsy samples with a minimum of six portal tracts) to those with low-quality liver biopsy (i.e. biopsy samples of a specified lower minimum number of portal tracts). Many studies did not report the minimum number of portal tracts required, so only two of the potential analyses were feasible: the Forns index low threshold for ruling out significant fibrosis (three versus seven studies, P = 0.692), and the FIB-4 low threshold for ruling out significant fibrosis (two versus three studies, P = 0.365). In these analyses, the quality of liver biopsy did not affect the results.

Studies with high versus low inflammation on liver biopsy

We planned to investigate the level of histological inflammation as a potential source of heterogeneity by comparing studies with a hepatitis activity index (HAI) of 0, 1, or 2 to studies with a mean HAI of 3 and 4. No studies specifically reported the level of histological inflammation and therefore this analysis was not possible.

Summary of findings

To summarise our findings, we focused on the condition for which each test was originally designed: the FIB-4 score for severe (≥ F3) fibrosis; and the Forns index for significant (≥ F2) fibrosis. These results are displayed in Summary of findings 1.

To diagnose/rule in severe fibrosis (≥ F3) using the high cut-off of 3.25, the FIB-4 score had a summary sensitivity of 41.4% (95% CI 33% to 50.4%) and summary specificity of 92.6% (95% CI 89.5% to 94.9%). By using the severe fibrosis median of 34.8%, a figure derived from the overall study cohort, we can expect the following results in a hypothetical cohort of 1000 people: 144 people with fibrosis test positive and are appropriately prioritised for antiviral treatment (true positives); 204 people with fibrosis test negative and are not prioritised for antiviral treatment (false negatives); 604 people without fibrosis test negative and are appropriately not prioritised for antiviral treatment (true negatives); and 48 people without fibrosis test positive and are inappropriately prioritised for antiviral treatment (false positives).

To exclude/rule out severe fibrosis (≥ F3) using the low cut-off of 1.45, the FIB-4 score had a summary sensitivity of 81.1% (95% CI 75.6% to 85.6%) and summary specificity of 62.3% (95% CI 57.4% to 66.9%). By using the severe fibrosis median of 30.9%, a figure derived from the overall study cohort, we can expect the following results in a hypothetical cohort of 1000 people: 251 people with fibrosis test positive and are appropriately prioritised for antiviral treatment (true positives); 58 people with fibrosis test negative and are not prioritised for antiviral treatment (false negatives); 430 people without fibrosis test negative and are appropriately not prioritised for antiviral treatment (true negatives); and 261 people without fibrosis test positive and are inappropriately prioritised for antiviral treatment (false positives).

To diagnose/rule in significant fibrosis (≥ F2) using the high cut-off of 6.9, the Forns index had a summary sensitivity of 34.1% (95% CI 26.4% to 42.8%) and summary specificity of 97.3% (95% CI 90.0% to 92.9%) for diagnosing significant (≥ F2) fibrosis. By using the significant fibrosis median of 52.5%, a figure derived from the overall study cohort, we can expect the following results in a hypothetical cohort of 1000 people: 179 people with fibrosis test positive and are appropriately prioritised for antiviral treatment (true positives); 346 people with fibrosis test negative and are not prioritised for antiviral treatment (false negatives); 462 people without fibrosis test negative and are appropriately not prioritised for antiviral treatment (true negatives); and 13 people without fibrosis test positive and are inappropriately prioritised for antiviral treatment (false positives).

To exclude/rule out significant fibrosis (≥ F2) using the low cut-off of 4.2, the Forns index had a summary sensitivity of 84.7% (95% CI 77.9% to 89.7%) and summary specificity of 47.9% (95% CI 38.6% to 57.3%). By using the significant fibrosis median of 54.5%, a figure derived from the overall study cohort, we can expect the following results in a hypothetical cohort of 1000 people: 462 people with fibrosis test positive and are appropriately prioritised for antiviral treatment (true positives); 83 people with fibrosis test negative and are not prioritised for antiviral treatment (false negatives); 218 people without fibrosis test negative and are appropriately not prioritised for antiviral treatment (true negatives); and 237 people without fibrosis test positive and are inappropriately prioritised for antiviral treatment (false positives).

DISCUSSION

Summary of main results

We aimed to determine the accuracy of the FIB-4 score and the Forns index for diagnosing three stages of hepatic fibrosis (F2 or worse, F3 or worse, and F4) in people with chronic hepatitis C compared with the reference standard of liver biopsy. We also attempted to compare the diagnostic accuracy of the FIB-4 score with that of the Forns index for specific fibrosis stages.

We identified 84 studies (107,583 participants) that met our inclusion criteria. In order to provide meaningful results for clinical practice, we restricted the meta-analysis to the 62 studies that reported cut-offs in a narrow range around the original validated cut-offs (+/- 0.15 for the FIB-4 score; +/- 0.3 for the Forns index).

Overall, we judged only two studies to have a low risk of bias in all domains, 23 studies (27%) to have an overall unclear risk of bias, and the remaining 59 studies (70%) to have an overall high risk of bias. Of the 84 included studies, we judged 13 (15%) to be of high concern for applicability, all due to participant selection processes. The single largest issue in the overall body of data in our review was authors deriving their cut-offs for the index test 'a posteriori' (instead of using a prespecified cut-off), which is likely to overestimate the diagnostic accuracy. Forty-three of the 59 (73%) studies with an overall high risk of bias did this, which made it our leading reason for rating these studies as having a high risk of bias overall.

FIB-4 score

The FIB-4 score was originally developed and validated for diagnosing severe fibrosis or worse (≥ F3). For this target condition, the summary sensitivity and specificity were 81.1% and 62.3%, respectively, for the low/rule-out cut-off, and 41.4% and 92.6% for the high/rule-in cut-off.

As described in the original publication (Sterling 2006), using the low cut-off of 1.45 allows the clinician to rule out severe fibrosis (≥ F3), whereas using the high cut-off of 3.25 allows the clinician to rule it in. When both cut-offs are used simultaneously, we calculated that 30.4% of people would have an indeterminate result, meaning a test reading higher than the low cut-off but lower than the high cut-off. This "grey area" represents people requiring further testing.

We also examined the FIB-4 score's performance in diagnosing significant fibrosis (F2) and cirrhosis (F4). The sensitivity and specificity results were similar to those for F3. As expected, the more severe the target condition (from F2 to F4) being assessed, the higher the sensitivity, and the lower the specificity.

According to our results, it is best to use the FIB-4 low cut-off of 1.45 to rule out severe fibrosis (≥ F3). In a hypothetical cohort of 1000 people (with a prevalence of severe fibrosis of 30.9%, according to our study), 430/488 (88%) of people testing negative would be correctly classified (i.e. fibrosis < F3). One would expect the value to be higher in an unselected population with a lower overall prevalence of severe fibrosis. Contrarily, when applying the high cut-off of 3.25, 48/192 (25%) of people testing positive would be wrongly diagnosed with fibrosis, suggesting that the FIB-4 score is relatively weak at accurately ruling in severe fibrosis.

Forns index

The Forns index was originally developed and validated for diagnosing significant fibrosis or worse (≥ F2). For this target condition, the summary sensitivity and specificity were 84.7% and 47.9%, respectively, for the low/rule-out cut-off, and 34.1% and 97.3% for the high/rule-in cut-off.

As described in the original publication (Forns 2002a), using the low cut-off of 4.2 allows the clinician to rule in severe fibrosis (≥ F3), whereas using the high cut-off of 6.9 allows the clinician to rule it out. When both cut-offs are used simultaneously, we calculated that 44.8% of people would have an indeterminate result, meaning a test reading higher than the low cut-off but lower than the high cut-off. This "grey area" represents people requiring further testing.

There were insufficient studies reporting on the Forns index's performance in diagnosing severe fibrosis or worse (≥ F3) and cirrhosis (F4) to confidently comment on its diagnostic accuracy in these target conditions.

According to our results, it is best to use the Forns index high cut-off of 6.9 to rule in significant fibrosis (≥ F2). In a hypothetical cohort of 1000 people (with a prevalence of significant fibrosis of 52.5%, according to our study), 179/192 (93%) of people testing positive would be correctly diagnosed. Contrarily, when applying the low cut-off of 4.2, 83/301 (28%) of people testing negative would be wrongly classified (i.e. would have significant fibrosis missed), suggesting that the Forns index is relatively weak at accurately ruling out significant fibrosis.

Comparing FIB-4 and Forns index

Due to the paucity of data, we could only compare the FIB-4 score to the Forns index at the low and high cut-offs for diagnosing significant fibrosis (≥ F2). For this target condition, there was no significant difference in the performance of these tests using either cut-off.

Comparison with previous research

The diagnostic performance of the FIB-4 score and the Forns index for liver fibrosis in people with chronic hepatitis C has not been extensively tested in systematic reviews or meta-analyses. The recent European Association for the Study of the Liver (EASL) Clinical Practice Guidelines on non-invasive tests for the evaluation of liver disease severity and prognosis only mentioned the accuracy of non-invasive tests in people who have already achieved sustained virological response, and did not provide guidelines for their use in people with active viral disease (EASL 2021).

Forns index

The low cut-off for ruling out significant fibrosis in people with chronic hepatitis C originally proposed by Forns and colleagues showed a high sensitivity of 94% (Forns 2002a). These results were later confirmed in 2013 when a meta-analysis found the Forns index low cut-off was able to rule out significant fibrosis with a sensitivity of 88% and an LR- of 0.22 (Chou 2013). The summary estimates from our meta-analysis were therefore slightly lower than these prior findings (sensitivity of 84.7%, LR- 0.32).

Macias and colleagues applied the low cut-off to a population of people with HIV and chronic hepatitis C infection and found that the

Forns index's ability to rule out significant fibrosis in this population was reduced (sensitivity 78%, Macías 2006). It has been proposed that this effect is related to confounding by hypercholesterolaemia, which is known to be more common in people living with HIV and can therefore affect the performance of the Forns index. In our meta-analysis, however, HIV status did not affect the performance of the index: there were no observable differences between studies with and without people living with HIV.

One relevant result of our meta-analysis is that the Forns index high cut-off seems able to rule in significant fibrosis (specificity 96%, LR + 12). This result is concordant with the results obtained in Forns 2002a and Macías 2006, where the high cut-off had a specificity of 95% and 96%, respectively. The meta-analysis by Chou and colleagues revealed a similar sensitivity of 94%, but an inferior LR + of 6.5. (Chou 2013).

In a meta-analysis of two studies on people with HIV and chronic hepatitis C infection, Shaheen and Myers proposed that the Forns index may have a role in excluding cirrhosis (Shaheen 2008). However, both our meta-analysis and the Chou 2013 meta-analysis show the Forns index high cut-off is insufficient in diagnosing cirrhosis (LR+ 7.4 and LR+ 5.4, respectively).

FIB-4

The ability of the FIB-4 score to rule out significant fibrosis (F2 or worse) found in our review was similar to that found in the Chou 2013 meta-analysis (sensitivity 76.2%, LR- 0.34 versus sensitivity 64%, LR- 0.53, respectively). However, compared to Chou 2013, we found an increased ability of the FIB-4 score at the high cut-off to rule in significant fibrosis (specificity 96.6%, LR+ 8.7 versus specificity 79%, LR+ 2.4, respectively).

Based on the results of our review, the FIB-4 low cut-off is a suitable test to rule out severe fibrosis or worse (sensitivity 81.1% and LR- 0.30). This exceeded the sensitivity of 66.7% found in the original paper by Sterling and colleagues, which was based on a cohort of people with chronic hepatitis C infection and HIV (Sterling 2006), and also exceeded the sensitivity of 74.3% found in a subsequent paper by Vallet-Pichard and colleagues in a cohort of people with chronic hepatitis C (Vallet-Pichard 2007).

The ability of the FIB-4 score to rule out cirrhosis found in this review was similar to that found in Chou 2013 (sensitivity 89.1%, LR- 0.20 versus sensitivity 90%, LR- 0.17, respectively). The findings of our review suggest that the FIB-4 score is a suitable non-invasive test to rule out cirrhosis, which is corroborated by Chou 2013. This finding could have implications for the clinical pathway in terms of the initiation of surveillance programmes for hepatocellular carcinoma and oesophageal varices.

Strengths and weaknesses of the review

Our review included 107,583 study participants from 28 countries over an 18-year period. This highlights the widespread use of these tests and the current relevance of the review question. The search strategy was inclusive and thorough. We included studies published in languages other than English and published as conference proceedings. We are confident we have missed few, if any, eligible studies, thereby reducing the chance of selection bias. Furthermore, by attempting to contact 32 authors for unpublished data, we allowed the appropriate inclusion or exclusion of an additional 17 papers, reducing the risk of data availability bias.

All major geographical areas were represented in our review apart from South America, sub-Saharan Africa, and Oceania. The FIB-4 score and the Forns index are potentially useful in resource-poor settings where access to patented non-invasive tests and liver biopsy is limited. Most data in our review came from resource-rich countries, and therefore the generalisability of our results to resource-poor settings is reduced. However, we do not think that there is a biological reason that these standardised laboratory tests would perform differently in a low-income setting, and therefore our results are most likely still useful for the clinical pathway in these settings.

Forty-three studies derived their optimal cut-offs 'a posteriori', consequently overestimating the diagnostic accuracy in the respective studies. Roughly half of these cut-offs were impossible to aggregate with the others to provide a summary estimate because they were beyond our limits of +/- 0.15 for the FIB-4 and +/- 0.3 for the Forns index (see Statistical analysis and data synthesis). To resolve this issue and to provide results in a clinically applicable manner, we restricted our analyses to narrow cut-off ranges, and excluded studies with unique and unmatched cut-offs. Whilst necessary, the exclusion of data from these 22 studies limits the overall power of the analyses.

When calculating the ‘grey area’ estimates, we restricted the analysis to data from studies that reported on both the FIB-4 score and the Forns index in the same research population (15 studies overall). Although unavoidable for this analysis, most studies reported either the FIB-4 score or the Forns index, and therefore, the power of these ‘grey area’ analyses is lower and the results less meaningful.

Several studies showed common methodological weaknesses, and according to our criteria, only two studies had an overall low risk of bias (Forns 2002a; Sebastiani 2008a), reducing the certainty of results from the overall body of evidence.

Since all data in this review come from referral centres, a further potential limitation is the ability to apply the results to the primary care population. The performance of the tests in the community might be different because the prevalence of advanced stages of fibrosis might be lower than that found in our review. In unselected populations, the probability of a false negative result decreases, significantly reducing the proportion of diseased people missed by the screening. On the other hand, such a prevalence change would also lead to an increase in the number of false positives which would then need to be ruled out in a referral centre.

Recent studies have raised the issue of age as a confounding factor for the FIB-4 score in various aetiologies of liver disease. Li and colleagues showed how the cut-offs, required to maintain an acceptable specificity when diagnosing various stages of fibrosis in people with chronic hepatitis B, differ depending on whether participants are younger or older than 30 years (Li 2017). Similarly, Ito and colleagues and McPherson and colleagues have explored the effects of age on the diagnostic accuracy of the FIB-4 score for various stages of liver fibrosis in the context of non-alcoholic fatty liver disease (Ito 2023; McPherson 2017). McPherson and colleagues showed that the specificity of the FIB-4 score for diagnosing severe fibrosis was unacceptably low in people older than 65 years. This confounding effect of age is an emerging area within the field and represents a further limitation to our review, since we did not have the necessary data at an individual patient level to perform a post

hoc analysis. The effect of age on non-invasive tests such as the FIB-4 score and the Forns index in the context of chronic hepatitis C would be an interesting area for future research.

We did not run some of our planned subgroup and sensitivity analyses, as fewer than four studies were available, meaning it was not possible to run the bivariate meta-analysis. This is an important limitation of the available evidence.

One final limitation concerns the heterogeneity analysis comparing studies with mean ALT values above 80 IU/L to those with mean ALT values at 80 IU/L or lower. These results should be interpreted with care, due to the risk of aggregation bias severely limiting the utility of the results.

Applicability of findings to the review question

Our primary objective was to determine the diagnostic test accuracy of the Forns index and the FIB-4 score in staging liver fibrosis in people with chronic hepatitis C virus, using liver biopsy as the reference standard. The target population was adults diagnosed with hepatitis C in any clinical setting (primary, secondary, or tertiary). We assessed applicability concerns by employing the three QUADAS-2 applicability domains: patient selection, index test, and reference standard (Figure 1; Figure 2; Figure 3). Applicability concerns were raised when the target population, index test, or reference standard of the included study did not suitably match our review question.

Overall, we judged 13 of the 84 included studies (15%) to be of high concern for applicability. Of the 74 studies that reported on the FIB-4 score, we deemed 11 (15%) to have high concerns for applicability. Of the 25 studies that reported on the Forns index, we deemed two (8%) to have high concerns for applicability. These 13 ‘high concern’ studies contributed 2735 participants (3%) to the meta-analysis, from the total of 100,605 participants in the 62 studies that reported cut-offs in a narrow range around the original validated cut-offs. Most participants contributing data to our summary estimates therefore came from studies with 'low concerns' for applicability.

With regard to participant selection, we judged 13 studies to be of high concern for applicability. In five studies, this was because the participant populations exclusively included a specific single subset of HCV genotypes (Bonnard 2015; Ferenci 2014; Gökan 2016; Shiha 2017; Usluer 2012). We judged four studies to be of high concern for applicability because the participant populations exclusively included liver transplant recipients (Corradi 2009; Kamphues 2010; Kitajima 2016; Segovia 2008), and three studies because the participant population exclusively included people with end-stage renal failure who were on dialysis or had received a renal transplant (Fontaine 2009; Patel 2017; Schmoyer 2020). We judged one study to be of high concern for applicability because the participant population was dominated by people with cryoglobulinemia (Sène 2006), which is known to be an independent risk factor for fibrosis (Saadoun 2006).

Our review question pertains to any setting – primary, secondary, or tertiary centres – as we are interested in the performance of the FIB-4 score and the Forns index as a triage tool. All studies included in this review were conducted in hospital settings, and none in primary care settings, potentially affecting the external generalisability of our results and the applicability of our summary estimates to primary care populations. In such populations, there would likely be a lower prevalence of fibrosis compared to patients in a hepatology clinic, and therefore the probability of a false negative result decreases significantly, reducing the proportion of diseased people missed by the screening; that is, the negative predictive value (NPV) in the general population is likely to be higher than our summary estimates. Such a prevalence change would also lead to an increase in the number of false positives, which would then need to be ruled out in a referral centre; that is, the positive predictive value (PPV) in the general population is likely to be lower than our summary estimates. However, in both low- and high-income countries, most people are triaged for management in a hospital setting. Furthermore, according to current guidelines, everyone with a hepatitis C viral infection should be assessed by a liver specialist (EASL 2020), and therefore we do not feel there are applicability issues with regard to the setting.

We judged all studies to have low concerns for applicability in terms of both the index test and the reference standard. This is because all included studies applied the FIB-4 score and the Forns index as described in the original publications (Forns 2002a; Sterling 2006). Similarly, in all included studies, liver biopsy was performed following suitable clinical guidelines and morphological results were estimated by a commonly used semi-quantitative score.

AUTHORS' CONCLUSIONS

Implications for practice

The European Association for the Study of the Liver (EASL) Clinical Practice Guidelines recommend hepatitis C antiviral treatment for everyone with a hepatitis C viral infection (EASL 2021). However, treatment remains restricted in many countries due to budget constraints. In these contexts, people with cirrhosis should be prioritised due to their increased risk of developing hepatocellular carcinoma and liver failure. Thus, the EASL Guidelines recommend a liver fibrosis assessment before the start of antiviral treatment (EASL 2021). In low- and middle-income countries, it is not feasible for such assessments to rely on expensive tests (such as FibroScan or magnetic resonance elastography) if there are cheaper, validated alternatives, such as the Fibrosis-4 (FIB-4) score and the Forns index.

Both the FIB-4 score and the Forns index may be considered for the initial assessment of people with chronic hepatitis C. The FIB-4 score's low cut-off (1.45) can be used to rule out people with at least severe fibrosis (≥ F3), and can be used to rule out those with cirrhosis (F4). The Forns index's high cut-off (6.9) can be used to diagnose people with at least significant fibrosis (≥ F2). Our review did not capture data from primary care populations. Thus, when generalising our results to an unselected primary care population, the probability of false positives will likely be higher and false negatives will likely be lower.

According to our data, approximately 30% and 50% of people will have an indeterminate result (i.e. their result is lower than the rule-in threshold and higher than the rule-out threshold) when the FIB-4 and Forns index tests are performed, respectively. For people in this 'grey area', a secondary test should be arranged to rule out liver fibrosis. The balance between the benefits of cheap and easy population screening and the costs of secondary testing to rule it out varies, depending on the local burden of disease, access

to tertiary care, and economic factors. These considerations were beyond the scope of our review.

Implications for research

Other studies addressing the diagnostic accuracy of the FIB-4 score and Forns index in hepatitis C as well as other liver conditions can refer to our results as a benchmark. Researchers investigating other non-invasive serological tests in the setting of hepatitis C may also find our summary estimates useful in their comparative research.

Data on the diagnostic accuracy of the FIB-4 score and Forns index in sub-Saharan Africa are missing, despite being a region that could particularly benefit from such inexpensive tests. Whatever the geographic region, more high-quality studies are needed: we judged most of the studies included in this review to be at unclear or high risk of bias.

The treatment landscape of hepatitis C has changed since the protocol for this Cochrane review was published in 2015. Today, it is recommended that everyone with chronic hepatitis C be treated with direct-acting antiviral medications, and most people achieve sustained virological response. It is uncertain what effect antiviral treatment has on the diagnostic accuracy of non-invasive tests such as the FIB-4 score and Forns index for liver fibrosis. Thus, studies, reviews, or both, that assess the diagnostic accuracy of non-invasive tests in people with chronic hepatitis C who have achieved sustained virological response represent an interesting area for future research.

ACKNOWLEDGEMENTS

Cochrane Review Group funding acknowledgement: the Danish State is the largest single funder of the Cochrane Hepato-Biliary Group through its investment in the Copenhagen Trial Unit, Centre for Clinical Intervention Research, the Capital Region of Denmark, Rigshospitalet, Copenhagen, Denmark. Disclaimer: the views and opinions expressed in this review are those of the authors and do not necessarily reflect those of the Danish State, the Capital Region, or the Copenhagen Trial Unit.

Cochrane Hepato-Biliary and Cochrane Diagnostic Test Accuracy Editorial Teams supported the authors in the development of this review.

The following people from the Cochrane Hepato-Biliary (CHB) Editorial Team conducted the editorial process for this article:

- Sign-off Editor (final editorial decision): Christian Gluud, Co-ordinating Editor CHBG, Denmark;
- Contact Editor (provided comments and editorial decision): Agostino Colli, Italy; Mirella Fraquelli, Italy;
- Managing Editor (selected peer reviewers, provided editorial guidance to authors, edited the article): Dimitrinka Nikolova, Denmark;
- Information Specialist (editorial guidance related to database searches): Sarah Louise Klingenberg, Denmark.

External peer review:

- Peer reviewer (provided clinical and content review comments): Ferenc Mozes, UK.

Cochrane Diagnostic Test Accuracy Editorial Team

- Sign-off Editor (DTAR methods): Ingrid Arevalo-Rodriguez, Screening and Diagnostic Tests Methods Group, UK
- Peer reviewers: Jérémie F. Cohen, UK (general methodology); Julie Glanville, UK (Information specialist); Nathan Pace (Statistician); Yemisi Takwoingi, University of Birmingham, UK (Statistician)
- Copy Editor (copy editing and production): Faith Armitage, Cochrane Central Production Service

REFERENCES

References to studies included in this review

Abdel-Hameed 2021a *{published and unpublished data}*

Abdel-Hameed EA, Rouster SD, Kottilil S, Sherman KE. The enhanced liver fibrosis index predicts hepatic fibrosis superior to FIB4 and APRI in HIV/HCV infected patients. *Clinical Infectious Diseases* 2021;**73**(3):450-9. [DOI: 10.1093/cid/ciaa646]

Abdel-Hameed 2021b *{published and unpublished data}*

Abdel-Hameed EA, Rouster SD, Kottilil S, Sherman KE. The enhanced liver fibrosis index predicts hepatic fibrosis superior to FIB4 and APRI in HIV/HCV infected patients. *Clinical Infectious Diseases* 2021;**73**(3):450-9. [DOI: 10.1093/cid/ciaa646]

Abdelsameea 2020 *{published and unpublished data}*

Abdelsameea E, Alsebaey A, Abdel-Razek W, Ehsan N, Morad W, Salama M, et al. Elastography and serum markers of fibrosis versus liver biopsy in 1270 Egyptian patients with hepatitis C. *European Journal of Gastroenterology & Hepatology* 2020;**32**(12):1553-8. [DOI: 10.1097/MEG.0000000000001672]

Abo El-Khair 2019 *{published data only}*

Abo El-Khair SM, El-Alfy HA, Elsamanoudy AZ, Elhammady D, Abd-Elfattah N, Eldeek B, et al. Development of a novel glycated protein-based fibrosis prediction score for determination of significant liver fibrosis in HCV-infected patients, a preliminary study. *Journal of Medical Virology* 2019;**92**(12):3525-33. [DOI: 10.1002/jmv.26204]

Ahmad 2011 *{published data only}*

Ahmad W, Ijaz B, Javed FT, Gull S, Kausar H, Sarwar MT, et al. A comparison of four fibrosis indexes in chronic HCV: development of new fibrosis-cirrhosis index (FCI). *BMC Gastroenterology* 2011;**11**:44. [DOI: 10.1186/1471-230X-11-44]

Alboraie 2015 *{published data only}*

Alboraie M, Khairy M, Elsharkawy M, Asem N, Elsharkawy A, Esmat G. Value of Egy-Score in diagnosis of significant, advanced hepatic fibrosis and cirrhosis compared to aspartate aminotransferase-to-platelet ratio index, FIB-4 and Forns' index in chronic hepatitis C virus. *Hepatology Research* 2015;**45**(5):560-70. [DOI: 10.1111/hepr.12385]

Amorim 2012 *{published data only}*

Amorim TG, Staub GJ, Lazzarotto C, Silva AP, Manes J, Ferronato Mda G, et al. Validation and comparison of simple noninvasive models for the prediction of liver fibrosis in chronic hepatitis C. *Annals of Hepatology* 2012;**11**(6):855-61.

Andrés-Otero 2016 *{published data only}*

Andrés-Otero MJ, De-Blas-Giral I, Puente-Lanzarote JJ, Serrano-Aulló T, Morandeira MJ, Lorente S, et al. Multiple approaches to assess fourteen non-invasive serum indexes for the diagnosis of liver fibrosis in chronic hepatitis C patients. *Clinical Biochemistry* 2016;**49**(7-8):560-5. [DOI: 10.1016/j.clinbiochem.2015.12.017]

Attallah 2012 *{published data only}*

Attallah AM, Omran MM, Farid K, El-Bendary M, Emran TM, Albannan MS, et al. Development of a novel score for liver fibrosis staging and comparison with eight simple laboratory scores in large numbers of HCV-monoinfected patients. *Clinica Chimica Acta* 2012;**413**(21-22):1725-30. [DOI: 10.1016/j.cca.2012.06.031]

Baldwin 2020 *{published data only}*

Baldwin N, Ravi S, Boudreaux J, Lin CP, Orandi B, Gray M. NALFD fibrosis score outperforms FIB-4 and Apri in predicting advanced fibrosis in non-NAFLD etiologies of liver disease. *Gastroenterology* 2020;**158**:1384-5. [0016-5085 1528-0012]

Bonnard 2015 *{published data only}*

Bonnard P, Elsharkawy A, Zalata K, Delarocque-Astagneau E, Biard L, Le Fouler L, et al. Comparison of liver biopsy and noninvasive techniques for liver fibrosis assessment in patients infected with HCV-genotype 4 in Egypt. *Journal of Viral Hepatitis* 2015;**22**(3):245-53. [DOI: 10.1111/jvh.12285]

Bourliere 2006 *{published data only}*

Bourliere M, Penaranda G, Renou C, Botta-Fridlund D, Tran A, Portal I, et al. Validation and comparison of indexes for fibrosis and cirrhosis prediction in chronic hepatitis C patients: proposal for a pragmatic approach classification without liver biopsies. *Journal of Viral Hepatitis* 2006;**3**(10):659-70. [DOI: 10.1111/j.1365-2893.2006.00736.x]

Calès 2010 *{published data only}*

Calès P, Halfon P, Batisse D, Carrat F, Perré P, Penaranda G, et al. Comparison of liver fibrosis blood tests developed for HCV with new specific tests in HIV/HCV co-infection. *Journal of Hepatology* 2010;**53**(2):238-44. [DOI: 10.1016/j.jhep.2010.03.007]

Cheng 2019 *{published data only}*

Cheng CH, Chu CY, Chen HL, Lin IT, Wu CH, Lee YK, et al. Subgroup analysis of the predictive ability of aspartate aminotransferase to platelet ratio index (APRI) and fibrosis-4 (FIB-4) for assessing hepatic fibrosis among patients with chronic hepatitis C. *Journal of Microbiology, Immunology & Infection* 2019;**53**(4):542-9. [DOI: 10.1016/j.jmii.2019.09.002]

Conti 2019 *{published data only}*

Conti F, Serra C, Vukotic R, Felicani C, Mazzotta E, Gitto S, et al. Assessment of liver fibrosis with elastography point quantification vs other noninvasive methods. *Clinical Gastroenterology & Hepatology* 2019;**17**(3):510-7.e3. [DOI: 10.1016/j.cgh.2018.06.027]

Cordie 2018 *{published data only}*

Cordie A, Salama A, El-Sharkawy M, El-Nahaas SM, Khairy M, Elsharkawy A, et al. Comparing the efficiency of Fib-4, Egy-score, APRI, and GUCI in liver fibrosis staging in Egyptians with chronic hepatitis C. *Journal of Medical Virology* 2018;**90**(6):1106-11. [DOI: 10.1002/jmv.25064]

Corradi 2009 *{published data only}*

Corradi F, Piscaglia F, Flori S, D'Errico-Grigioni A, Vasuri F, Tame MR, et al. Assessment of liver fibrosis in transplant recipients with recurrent HCV infection: usefulness of transient elastography. *Digestive and Liver Disease* 2009;**41**(3):217-25.

Crisan 2012 *{published data only}*

Crisan D, Radu C, Lupsor M, Sparchez Z, Grigorescu MD, Grigorescu M. Two or more synchronous combination of noninvasive tests to increase accuracy of liver fibrosis assessement in chronic hepatitis C; results from a cohort of 446 patients. *Hepatitis monthly* 2012;**12**(3):177. [DOI: 10.5812/hepatmon.853]

Demma 2018 *{published and unpublished data}*

Demma S, Benmassaoud A, Campbell S, Macdonald D. An evaluation of Elift, Fib-4 and APRI for exclusion of advanced fibrosis in a real-world Hepatitis C treatment cohort. *Hepatology* 2018;**68**(Suppl 1):910A-1A.

De Oliveira 2016 *{published data only}*

De Oliveira AC, El-Bacha I, Vianna MV, Parise ER. Utility and limitations of APRI and FIB4 to predict staging in a cohort of nonselected outpatients with hepatitis C. *Annals of Hepatology* 2016;**15**(3):326-32. [DOI: 10.5604/16652681.1198801]

Ferenci 2014 *{published and unpublished data}*

Ferenci P, Aires R, Beavers KL, Curescu M, Abrao Ferreira PR, Gschwantler M, et al. Predictive value of FIB-4 and APRI versus METAVIR on sustained virologic response in genotype 1 hepatitis C patients. *Hepatology International* 2014;**8**(1):83-93. [DOI: 10.1007/s12072-013-9484-6]

Fontaine 2009 *{published data only}*

Fontaine H, Nalpas B, Vallet-Pichard A, Mallet V, Pol S. Low diagnosis accuracy of fibrotest and FIB-4 in HCV-infected haemodialysis and kidney recipients. *Hepatology* 2009;**50**(Suppl 4):1064A.

Forns 2002a *{published data only}*

Forns X, Ampurdanes S, Llovet JM, Aponte J, Quintó L, Martınez-Bauer E, et al. Identification of chronic hepatitis C patients without hepatic fibrosis by a simple predictive model. *Hepatology* 2002;**36**(4):986-92. [DOI: 10.1053/jhep.2002.36128]

Forns 2002b *{published data only}*

Forns X, Ampurdanes S, Llovet JM, Aponte J, Quintó L, Martınez-Bauer E, et al. Identification of chronic hepatitis C patients without hepatic fibrosis by a simple predictive model. *Hepatology* 2002;**36**(4):986-92. [DOI: 10.1053/jhep.2002.36128]

Fouad 2018 *{published data only}*

Fouad R, Elbaz T, Abdel Alem S, Elsharkawy A, Negm M, Khairy M, et al. Evaluation of accuracy of elastography point quantification versus other noninvasive modalities in staging of fibrosis in chronic hepatitis C virus patients. *European Journal of Gastroenterology and Hepatology* 2018;**30**:882-7.

Fujita 2018 *{published data only}*

Fujita K, Kuroda N, Morishita A, Oura K, Tadokoro T, Nomura T, et al. Fibrosis staging using direct serum biomarkers is influenced by hepatitis activity grading in hepatitis C virus infection. *Journal of Clinical Medicine* 2018;**7**(9):11.

Gamil 2017 *{published data only}*

Gamil M, Alboraie M, El-Sayed M, Elsharkawy A, Asem N, Elbaz T, et al. Novel scores combining AFP with non-invasive markers for prediction of liver fibrosis in chronic hepatitis C patients. *Journal of Medical Virology* 2017;**90**(6):1080-6.

Gökan 2016 *{published data only}*

Gökcan H, Kuzu UB, Öztaş E, Saygılı F, Öztuna D, Suna N, et al. The predictive value of noninvasive serum markers of liver fibrosis in patients with chronic hepatitis C. *Turkish Journal of Gastroenterology* 2016;**27**(2):156-64. [DOI: 10.5152/tjg.2015.150449]

Gorka-Dynysiewicz 2019 *{published data only}*

Gorka-Dynysiewicz J, Pazgan-Simon M, Zuwala-Jagiello J. Pentraxin 3 detects clinically significant fibrosis in patients with chronic viral hepatitis C. *BioMed Research International* 2019;**2019**:2639248. [DOI: 10.1155/2019/2639248]

Gorka-Dynysiewicz 2020 *{published data only}*

Gorka-Dynysiewicz J, Pazgan-Simon M, Zuwala-Jagiello J. Accurate prediction of significant liver fibrosis using the Pentra score model in patients with chronic hepatitis C. *Polish Archives Of Internal Medicine* 2020;**130**(2):112-20.

Gounder 2018 *{published data only}*

Gounder PP, Haering C, Bruden DJ, Townshend-Bulson L, Simons BC, Spradling PR, et al. Does incorporating change in APRI or FIB-4 indices over time improve the accuracy of a single index for identifying liver fibrosis in persons with chronic hepatitis C virus infection? *Journal of Clinical Gastroenterology* 2018;**52**(1):60-6. [DOI: 10.1097/MCG.0000000000000753]

Guilabert 2010 *{published data only}*

Guilabert MI, Mena-Bernal CH, del Pozo González J, del Pozo Pérez MA. Retrospective diagnostic accuracy study of liver fibrosis of Fibroscan, APRI, FIB-4 and Forns compared to biopsy in HCV patients both with and without co-infection of HIV [Estudio retrospectivo de la capacidad de evaluación de fibrosis hepática del FibroScan®, APRI, FIB-4 y FORNS con referencia a la biopsia hepática de pacientes con hepatitis crónica C, mono y coinfectados con VIH]. *Gastroenterología y Hepatología* 2010;**33**(6):425-32. [DOI: 10.1016/j.gastrohep.2010.02.005]

Güzelbulut 2011 *{published data only}*

Güzelbulut F, Akkan Cetinkya Z, Sezikli M, Yaşar B, Özkara S, Kurdaş Övünç AO. AST-platelet ratio index, Forns index and FIB-4 in the prediction of significant fibrosis and cirrhosis in patients with chronic hepatitis C [Kronik hepatit C'li hastalarda belirgin fibrozis ve sirozun belirlenmesinde AST-platelet oran› indeksi, Forns indeksi ve FIB-4'ün yeri]. *Turkish Journal of Gastroenterolgy* 2011;**22**(3):279-85. [DOI: 10.4318/tjg.2011.0213]

Hashem 2021 *{published data only}*

Hashem A, Awad A, Shousha H, Alakel W, Salama A, Awad T, et al. Validation of a machine learning approach using FIB-4 and APRI scores assessed by the metavir scoring system: a cohort

study. *Arab Journal of Gastroenterology* 2021;**22**(2):88-92. [DOI: 10.1016/j.ajg.2021.04.003]

Holmberg 2013 *{published data only}*

Holmberg SD, Lu M, Rupp LB, Lamerato LE, Moorman AC, Vijayadeva V, et al. Noninvasive serum fibrosis markers for screening and staging chronic hepatitis C virus patients in a large US cohort. *Clinical Infectious Diseases* 2013;**57**(2):240-6. [DOI: 10.1093/cid/cit245]

Hseih 2012 *{published data only}*

Hsieh YY, Tung SY, Lee K, Wu CS, Wei KL, Shen CH, et al. Routine blood tests to predict liver fibrosis in chronic hepatitis C. *World Journal of Gastroenterology* 2012;**18**(8):746. [PMID: 22371634]

Hsu 2019 *{published and unpublished data}*

Hsu TH, Tsui PH, Yu WT, Huang SF, Tai J, Wan YL, et al. Cutoff values of acoustic radiation force impulse two-location measurements in different etiologies of liver fibrosis. *Journal of Medical Ultrasound* 2019;**27**(3):130-4. [DOI: 10.4103/JMU.JMU_7_19]

Iacobellis 2005 *{published data only}*

Iacobellis A, Mangia A, Leandro G, Clemente R, Festa V, Attino V, et al. External validation of biochemical indices for noninvasive evaluation of liver fibrosis in HCV chronic hepatitis. *American College of Gastroenterology* 2005;**100**(4):868-73. [DOI: 10.1111/j.1572-0241.2005.40881.x]

Ikatura 2021 *{published and unpublished data}*

Itakura J, Kurosaki M, Setoyama H, Simakami T, Oza N, Korenaga M, et al. Applicability of APRI and FIB-4 as a transition indicator of liver fibrosis in patients with chronic viral hepatitis. *Journal of Gastroenterology* 2021;**56**(5):470-8. [DOI: 10.1007/s00535-021-01782-3]

Kamphues 2010 *{published data only}*

Kamphues C, Lotz K, Röcken C, Berg T, Eurich D, Pratschke J, et al. Chances and limitations of non-invasive tests in the assessment of liver fibrosis in liver transplant patients. *Clinical Transplantation* 2010;**24**(5):652-9. [DOI: 10.1111/j.1399-0012.2009.01152.x]

Kitajima 2016 *{published data only}*

Kitajima T, Kaido T, Hamaguchi Y, Yagi S, Taura K, Fujimoto Y, et al. Validation of the FIB-4 index for evaluation of fibrosis in patients with recurrent hepatitis C after living donor liver transplantation: a single center experience. *Hepatology Research* 2016;**46**(8):752-7. [DOI: 10.1111/hepr.12617]

Koller 2014 *{published data only}*

Koller T, Kollerova J, Huorka M, Meciarova I, Payer J. Noninvasive scoring algorithm to identify significant liver fibrosis among treatment-naive chronic hepatitis C patients. *European Journal of Gastroenterology & Hepatology* 2014;**26**(10):1108-15. [DOI: 10.1097/MEG.0000000000000182]

Ladero 2010 *{published data only}*

Ladero JM, Delkader J, Ortega L, Fernández C, Devesa MJ, López-Alonso G, et al. Non-invasive evaluation of the fibrosis stage in chronic hepatitis C: a comparative analysis of nine scoring methods. *Scandinavian Journal of Gastroenterology* 2010;**45**(1):51-9. [DOI: 10.3109/00365520903305544]

Leroy 2007 *{published data only}*

Leroy V, Hilleret MN, Sturm N, Trocme C, Renversez JC, Faure P, et al. Prospective comparison of six non-invasive scores for the diagnosis of liver fibrosis in chronic hepatitis C. *Journal of Hepatology* 2007;**46**(5):775-82. [DOI: 10.1016/j.jhep.2006.12.013]

Loko 2008 *{published data only}*

Loko MA, Castera L, Dabis F, Le Bail B, Winnock M, Coureau G, et al, Groupe d'Epidémiologie Clinique du SIDA en Aquitaine. Validation and comparison of simple noninvasive indexes for predicting liver fibrosis in HIV-HCV-coinfected patients: ANRS CO3 Aquitaine cohort. *American Journal of Gastroenterology* 2008;**103**(8):1973-80. [DOI: 10.1111/j.1572-0241.2008.01954.x]

Macías 2006 *{published data only}*

Macías J, Girón-González JA, González-Serrano M, Merino D, Cano P, Mira JA, et al. Prediction of liver fibrosis in human immunodeficiency virus/hepatitis C virus coinfected patients by simple non-invasive indexes. *Gut* 2006;**55**(3):409-14. [DOI: 10.1136/gut.2005.065904]

Maheshwari 2013 *{published data only}*

Maheshwari R, Yang J, Miller L. FIB-4 Index as a non-invasive and inexpensive predictor of hepatic fibrosis in a primarily uninsured African-American population with chronic hepatitis C. *American Journal of Gastroenterology* 2013;**108**(Suppl 1):S112. [DOI: 10.1038/ajg.2013.265]

Martinez 2011 *{published and unpublished data}*

Martinez SM, Fernández-Varo G, González P, Sampson E, Bruguera M, Navasa M, et al. Assessment of liver fibrosis before and after antiviral therapy by different serum marker panels in patients with chronic hepatitis C. *Alimentary Pharmacology & Therapeutics* 2011;**33**(1):138-48. [DOI: 10.1111/j.1365-2036.2010.04500.x]

Matsuura 2018 *{published data only}*

Matsuura K, Aizawa N, Enomoto H, Nishiguchi S, Toyoda H, Kumada T, et al. Circulating let-7 Levels in Serum Correlate With the Severity of Hepatic Fibrosis in Chronic Hepatitis C. *Open Forum Infectious Diseases* 2018;**5**(11):268. [DOI: 10.1093/ofid/ofy268]

Mossong 2011 *{published data only}*

Mossong J, Bill S, Hawotte K, Gilson G, Knolle U, Weber J, et al. Predicting significant fibrosis in hepatitis C patients in Luxembourg using serological markers. *Bulletin de la Société des Sciences Médicales du Grand-Duché de Luxembourg* 2011;**1**:19-30.

Nan 2019 *{published data only}*

Nan Y, Niu X, Wang R, Zhao S, Fu N, Du J, et al. microRNA-1273g-3p is a useful non-invasive test for the prediction of liver fibrosis in patients with chronic hepatitis C. *Experimental & Therapeutic Medicine* 2019;**17**(3):1817-24.

Omran 2018 *{published data only}*

Omran D, Zayed RA, Nabeel MM, Mobarak L, Zakaria Z, Farid A, et al. Evaluating diagnostic accuracy of noninvasive tests in assessment of significant liver fibrosis in chronic hepatitis C Egyptian patients. *Viral Immunology* 2018;**31**(4):315-20. [DOI: 10.1089/vim.2017.0134]

Ozel 2015 *{published data only}*

Ozel BD, Poyrazoğlu OK, Karaman A, Karaman H, Altinkaya E, Sevinç E, et al. The PAPAS index: a novel index for the prediction of hepatitis C-related fibrosis. *European Journal of Gastroenterology & Hepatology* 2015;**27**(8):895-900. [DOI: 10.1097/MEG.0000000000000379]

Paranaguá-Vezozzo 2017 *{published data only}*

Paranaguá-Vezozzo DC, Andrade A, Mazo DF, Nunes V, Guedes AL, Ragazzo TG, et al. Concordance of non-invasive mechanical and serum tests for liver fibrosis evaluation in chronic hepatitis C. *World Journal of Hepatology* 2017;**9**(8):436. [DOI: 10.4254/wjh.v9.i8.436]

Patel 2017 *{published data only}*

Patel R, Rassameehiran S, Patil R, Dhingra S, Sood GK. Diagnostic utility of non-invasive markers for staging of liver disease in hepatitis C patients with ESRD on haemodialysis. *Gastroenterology* 2017;**152**(5):S1165-6.

Portilla 2009 *{published data only}*

Portilla J, López-Burgos A, Saiz-De-La-Hoya-Zamácola P, Sánchez-Payá J, Bedía-Collantes M, Faraco-Atienzar I, et al. Utility of 2 predictive biochemical models of hepatic fibrosis grade in the prison population with hepatitis C [Utilidad de 2 modelos bioquímicos predictivos del grado de fibrosis hepática en la población penitenciaria con hepatitis C]. *Gastroenterología y Hepatología* 2009;**32**(6):387-94. [DOI: 10.1016/j.gastrohep.2009.01.176]

Qian 2019 *{published data only}*

Qian X, Zheng S, Wang L, Yao M, Guan G, Wen X, et al. Exploring the diagnostic potential of serum golgi protein 73 for hepatic necroinflammation and fibrosis in chronic HCV infection with different stages of liver injuries. *Disease Markers* 2019;**2019**:3862024. [DOI: 10.1155/2019/3862024]

Ramzy 2021 *{published data only}*

Ramzy I, Fouad R, Salama R, Abdellatif Z, Elsharkawy A, Zayed N, et al. Evaluation of red cell distribution width to platelet ratio as a novel non-invasive index for predicting hepatic fibrosis in patients with chronic hepatitis C. *Arab Journal of Gastroenterology* 2021;**22**(1):6-11. [DOI: 10.1016/j.ajg.2020.12.003]

Schmid 2015 *{published data only}*

Schmid P, Bregenzer A, Huber M, Rauch A, Jochum W, Müllhaupt B, et al, Swiss HIV Cohort Study. Progression of liver fibrosis in HIV/HCV co-infection: a comparison between non-invasive assessment methods and liver biopsy. *PLOS One* 2015;**10**(9):e0138838. [DOI: 10.1371/journal.pone.0138838]

Schmoyer 2020 *{published data only}*

Schmoyer CJ, Kumar D, Gupta G, Sterling RK. Diagnostic accuracy of noninvasive tests to detect advanced hepatic fibrosis in patients with hepatitis C and end-stage renal disease. *Clinical Gastroenterology & Hepatology* 2020;**18**(10):2332-9. [DOI: 10.1016/j.cgh.2020.02.019]

Sebastiani 2008a *{published data only}*

Sebastiani G, Vario A, Guido M, Alberti A. Performance of noninvasive markers for liver fibrosis is reduced in chronic hepatitis C with normal transaminases. *Journal of Viral Hepatitis* 2008;**15**(3):212-8. [DOI: 10.1111/j.1365-2893.2007.00932.x]

Sebastiani 2008b *{published data only}*

Sebastiani G, Vario A, Guido M, Alberti A. Performance of noninvasive markers for liver fibrosis is reduced in chronic hepatitis C with normal transaminases. *Journal of Viral Hepatitis* 2008;**15**(3):212-8. [DOI: 10.1111/j.1365-2893.2007.00932.x]

Sebastiani 2012 *{published data only}*

Sebastiani G, Halfon P, Castera L, Mangia A, Di Marco V, Pirisi M, et al. Comparison of three algorithms of non-invasive markers of fibrosis in chronic hepatitis C. *Alimentary Pharmacology & Therapeutics* 2012;**35**(1):92-104. [DOI: 10.1111/j.1365-2036.2011.04897.x]

Segovia 2008 *{published data only}*

Segovia MC, Lyden ER, Bernard T, McCashland TM. Evaluation of FIB-4 as a marker of fibrosis in HCV infected patients who underwent liver transplantation. *Hepatology* 2008;**48**(4 Suppl):576A.

Sène 2006 *{published data only}*

Sène D, Limal N, Messous D, Ghillani-Dalbin P, Charlotte F, Thiollière JM, et al. Biological markers of liver fibrosis and activity as non-invasive alternatives to liver biopsy in patients with chronic hepatitis C and associated mixed cryoglobulinemia vasculitis. *Clinical Biochemistry* 2006;**39**(7):715-21. [DOI: 10.1016/j.clinbiochem.2006.04.019]

Shaikh 2009 *{published data only}*

Shaikh S, Memon MS, Ghani H, Baloch GH, Jaffery M, Shaikh K. Validation of three non-invasive markers in assessing the severity of liver fibrosis in chronic hepatitis C. *Journal of the College of Physicians and Surgeons Pakistan* 2009;**19**(8):478-82.

Shiha 2017 *{published data only}*

Shiha G, Seif S, Eldesoky A, Elbasiony M, Soliman R, Metwally A, et al. A simple bedside blood test (Fibrofast; FIB-5) is superior to FIB-4 index for the differentiation between non-significant and significant fibrosis in patients with chronic hepatitis C. *Hepatology International* 2017;**11**(3):286-91. [DOI: 10.1007/s12072-017-9796-z]

Shiha 2022a *{published and unpublished data}*

Shiha G, Soliman R, Mikhail NN, Alswat K, Abdo A, Sanai F, et al. Development and multicenter validation of FIB-6: a novel, machine learning, simple bedside score to rule out liver cirrhosis and compensated advanced chronic liver disease in patients with chronic hepatitis C. *Hepatology Research* 2022;**52**:165-75. [DOI: 10.1111/hepr.13729]

Shiha 2022b *{published and unpublished data}*

Shiha G, Soliman R, Mikhail NN, Alswat K, Abdo A, Sanai F, et al. Development and multicenter validation of FIB-6: A novel, machine learning, simple bedside score to rule out liver cirrhosis and compensated advanced chronic liver disease in patients with chronic hepatitis C. *Hepatology Research* 2022;**52**:165-75. [DOI: 10.1111/hepr.13729]

Shiha 2022c *{published and unpublished data}*

Shiha G, Soliman R, Mikhail NN, Alswat K, Abdo A, Sanai F, et al. Development and multicenter validation of FIB-6: A novel, machine learning, simple bedside score to rule out liver cirrhosis and compensated advanced chronic liver disease in patients with chronic hepatitis C. *Hepatology Research* 2022;**52**:165-75. [DOI: 10.1111/hepr.13729]

Shiha 2022d *{published and unpublished data}*

Shiha G, Soliman R, Mikhail NN, Alswat K, Abdo A, Sanai F, et al. Development and multicenter validation of FIB-6: A novel, machine learning, simple bedside score to rule out liver cirrhosis and compensated advanced chronic liver disease in patients with chronic hepatitis C. *Hepatology Research* 2022;**52**:165-75. [DOI: 10.1111/hepr.13729]

Shiha 2022e *{published and unpublished data}*

Shiha G, Soliman R, Mikhail NN, Alswat K, Abdo A, Sanai F, et al. Development and multicenter validation of FIB-6: A novel, machine learning, simple bedside score to rule out liver cirrhosis and compensated advanced chronic liver disease in patients with chronic hepatitis C. *Hepatology Research* 2022;**52**:165-75. [DOI: 10.1111/hepr.13729]

Shiha 2022f *{published and unpublished data}*

Shiha G, Soliman R, Mikhail NN, Alswat K, Abdo A, Sanai F, et al. Development and multicenter validation of FIB-6: A novel, machine learning, simple bedside score to rule out liver cirrhosis and compensated advanced chronic liver disease in patients with chronic hepatitis C. *Hepatology Research* 2022;**52**:165-75. [DOI: 10.1111/hepr.13729]

Shiha 2022g *{published and unpublished data}*

Shiha G, Soliman R, Mikhail NN, Alswat K, Abdo A, Sanai F, et al. Development and multicenter validation of FIB-6: A novel, machine learning, simple bedside score to rule out liver cirrhosis and compensated advanced chronic liver disease in patients with chronic hepatitis C. *Hepatology Research* 2022;**52**:165-75. [DOI: 10.1111/hepr.13729]

Shiha 2022h *{published and unpublished data}*

Shiha G, Soliman R, Mikhail NN, Alswat K, Abdo A, Sanai F, et al. Development and multicenter validation of FIB-6: A novel, machine learning, simple bedside score to rule out liver cirrhosis and compensated advanced chronic liver disease in patients with chronic hepatitis C. *Hepatology Research* 2022;**52**:165-75. [DOI: 10.1111/hepr.13729]

Shiha 2022i *{published and unpublished data}*

Shiha G, Soliman R, Mikhail NN, Alswat K, Abdo A, Sanai F, et al. Development and multicenter validation of FIB-6: A novel, machine learning, simple bedside score to rule out liver cirrhosis and compensated advanced chronic liver disease in patients with chronic hepatitis C. *Hepatology Research* 2022;**52**:165-75. [DOI: 10.1111/hepr.13729]

Shiha 2022j *{published and unpublished data}*

Shiha G, Soliman R, Mikhail NN, Alswat K, Abdo A, Sanai F, et al. Development and multicenter validation of FIB-6: A novel, machine learning, simple bedside score to rule out liver cirrhosis and compensated advanced chronic liver disease in patients with chronic hepatitis C. *Hepatology Research* 2022;**52**:165-75. [DOI: 10.1111/hepr.13729]

Silva Junior 2014 *{published data only}*

Silva Junior RG, Schmillevitch J, Nascimento MD, Miranda ML, Brant PE, Schulz PO, et al. Acoustic radiation force impulse elastography and serum fibrosis markers in chronic hepatitis C. *Scandinavian Journal of Gastroenterology* 2014;**49**(8):986-92. [DOI: 10.3109/00365521.2014.909528]

Şirli 2010 *{published data only}*

Şirli R, Sporea I, Bota S, Popescu A, Cornianu M. A comparative study of non-invasive methods for fibrosis assessment in chronic HCV infection. *Hepatitis Monthly* 2010;**10**(2):88-94.

Stauber 2015 *{published data only}*

Stauber RE, Maieron A, Posch A, Pock H, Streit A, Lackner C. FIB-4, a simple fibrosis test, accurately predicts METAVIR F3/F4 stage in chronic hepatitis C. *United European Gastroenterology Journal* 2015;**3**(5S):A519.

Sterling 2006 *{published and unpublished data}*

Sterling RK, Lissen E, Clumeck N, Sola R, Correa MC, Montaner J, et al. Development of a simple noninvasive index to predict significant fibrosis in patients with HIV/HCV coinfection. *Hepatology* 2006;**43**(6):1317-25. [DOI: 10.1002/hep.21178]

Stibbe 2011 *{published data only}*

Stibbe KJ, Verveer C, Francke J, Hansen BE, Zondervan PE, Kuipers EJ, et al. Comparison of non-invasive assessment to diagnose liver fibrosis in chronic hepatitis B and C patients. *Scandinavian Journal of Gastroenterology* 2011;**46**(7-8):962-72. [DOI: 10.3109/00365521.2011.574725]

Tachi 2015 *{published data only}*

Tachi Y, Hirai T, Toyoda H, Tada T, Hayashi K, Honda T, et al. Predictive ability of laboratory indices for liver fibrosis in patients with chronic hepatitis C after the eradication of hepatitis C virus. *PLoS One* 2015;**10**(7):e0133515. [DOI: 10.1371/journal.pone.0133515]

Tanwar 2017 *{published data only}*

Tanwar S, Trembling PM, Hogan BJ, Parkes J, Harris S, Grant P, et al. Biomarkers of hepatic fibrosis in chronic hepatitis C. *Journal of Clinical Gastroenterology* 2017;**51**(3):268-77. [DOI: 10.1097/MCG.0000000000000581]

Toson 2017 *{published data only}*

Toson ES, Shiha GE, El-Mezayen HA, Samir W, El-Khininy MM. Noninvasive estimation of liver fibrosis in biopsy-proven hepatitis C virus-infected patients: angiogenic fibrogenic link. *European Journal of Gastroenterology & Hepatology* 2017;**29**(2):199-207. [DOI: 10.1097/MEG.0000000000000775]

Trang 2008 *{published data only}*

Trang T, Petersen JR, Snyder N. Non-invasive markers of hepatic fibrosis in patients co-infected with HCV and HIV: comparison of the APRI and FIB-4 index. *Clinica Chimica Acta* 2008;**397**(1-2):51-4. [DOI: 10.1016/j.cca.2008.07.009]

Trifan 2009 *{published data only}*

Trifan A, Cojocariu C, Sfarti C, Stanciu C. Prospective clinical trial on the accuracy of non-invasive tests to predict liver fibrosis in chronic HCV patients. *Hepatology* 2009;**50**(4 Suppl):1074A-5A.

Tsukano 2017 *{published data only}*

Tsukano N, Miyase S, Saeki T, Mizobe K, Iwashita H, Arima N, et al. Usefulness of virtual touch quantification for staging liver fibrosis in patients with hepatitis C, and factors affecting liver stiffness measurement failure compared with liver biopsy. *Hepatology Research* 2018;**48**(5):373-82. [DOI: 10.1111/hepr.13041]

Tural 2009 *{published data only}*

Tural C, Tor J, Sanvisens A, Pérez–Alvarez N, Martínez E, Ojanguren I, et al. Accuracy of simple biochemical tests in identifying liver fibrosis in patients co-infected with human immunodeficiency virus and hepatitis C virus. *Clinical Gastroenterology and Hepatology* 2009;**7**(3):339-45. [DOI: 10.1016/j.cgh.2008.11.019]

Udompap 2020 *{published data only}*

Udompap P, Sukonrut K, Suvannarerg V, Pongpaibul A, Charatcharoenwitthaya P. Prospective comparison of transient elastography, point shear wave elastography, APRI and FIB-4 for staging liver fibrosis in chronic viral hepatitis. *Journal of Viral Hepatitis* 2020;**27**(4):437-48. [DOI: 10.1111/jvh.13246]

Usluer 2012 *{published data only}*

Usluer GA, Erben NU, Aykin N, Dagli O, Aydogdu O, Barut S, et al. Comparison of non-invasive fibrosis markers and classical liver biopsy in chronic hepatitis C. *European Journal of Clinical Microbiology & Infectious Diseases* 2012;**31**(8):1873-8. [DOI: 10.1007/s10096-011-1513-6]

Vallet-Pichard 2007 *{published data only}*

Vallet-Pichard A, Mallet V, Nalpas B, Verkarre V, Nalpas A, Dhalluin-Venier V, et al. FIB-4: an inexpensive and accurate marker of fibrosis in HCV infection. Comparison with liver biopsy and fibrotest. *Hepatology* 2007;**46**(1):32-6. [DOI: 10.1002/hep.21669]

Wang 2015 *{published data only}*

Wang CC, Liu CH, Lin CL, Wang PC, Tseng TC, Lin HH, et al. Fibrosis index based on four factors better predicts advanced fibrosis or cirrhosis than aspartate aminotransferase/platelet ratio index in chronic hepatitis C patients. *Journal of the Formosan Medical Association* 2015;**114**(10):923-8. [DOI: 10.1016/j.jfma.2015.07.004]

Wang 2017 *{published data only}*

Wang HW, Peng CY, Lai HC, Su WP, Lin CH, Chuang PH, et al. New noninvasive index for predicting liver fibrosis in Asian patients with chronic viral hepatitis. *Scientific Reports* 2017;**7**(1):1-8. [DOI: 10.1038/s41598-017-03589-w]

Yen 2018 *{published data only}*

Yen YH, Kuo FY, Kee KM, Chang KC, Tsai MC, Hu TH, et al. APRI and FIB-4 in the evaluation of liver fibrosis in chronic hepatitis C patients stratified by AST level. *PLoS One* 2018;**13**(6):e0199760. [DOI: 10.1371/journal.pone.0199760]

Yilmaz 2021 *{published data only}*

Yilmaz B, Kayadibi H, Yeniova AO, Koseoglu H, Simsek Z. The age, bilirubin and albumin (ABA) index: a novel noninvasive index for predicting liver fibrosis in patients with chronic hepatitis C infection. *European Journal of Gastroenterology & Hepatology* 2021;**33**(1):e290-6.

Additional references

Bedossa 2003

Bedossa P, Dargere D, Paradis V. Sampling variability of liver fibrosis in chronic hepatitis C. *Hepatology* 2003;**38**(6):1449-57.

Bonekamp 2009

Bonekamp S, Kamel I, Solga S, Clark J. Can imaging modalities diagnose and stage hepatic fibrosis and cirrhosis accurately? *Journal of Hepatology* 2009;**50**(1):17-35.

Bravo 2001

Bravo AA, Sheth SG, Chopra S. Liver biopsy. *New England Journal of Medicine* 2001;**344**(7):495-500.

Castera 2008

Castera L, Forns X, Alberti A. Non-invasive evaluation of liver fibrosis using transient elastography. *Journal of Hepatology* 2008;**48**(5):835-47.

Castera 2009

Castera L. Transient elastography and other noninvasive tests to assess hepatic fibrosis in patients with viral hepatitis. *Journal of Viral Hepatitis* 2009;**16**(5):300-14.

Cholongitas 2006

Cholongitas E, Senzolo M, Standish R, Marelli L, Quaglia A, Patch D, et al. A systematic review of the quality of liver biopsy specimens. *American Journal of Clinical Pathology* 2006;**125**(5):710-21.

Chou 2013

Chou R, Wasson N. Blood tests to diagnose fibrosis or cirrhosis in patients with chronic hepatitis C virus infection: a systematic review. *Annals of Internal Medicine* 2013;**158**(11):807-20.

Colloredo 2003

Colloredo G, Guido M, Sonzogni A, Leandro G. Impact of liver biopsy size on histological evaluation of chronic viral hepatitis: the smaller the sample, the milder the disease. *Journal of Hepatology* 2003;**39**(2):239-44.

Crossan 2015

Crossan C, Tsochatzis EA, Longworth L, Gurusamy K, Davidson B, Rodríguez-Perálvarez M, et al. Cost-effectiveness of non-invasive methods for assessment and monitoring of liver fibrosis and cirrhosis in patients with chronic liver disease:

systematic review and economic evaluation. *Health Technology Assessment* 2015;**19**(9):1-409.

Deeks 2023

Deeks JJ, Bossuyt PM, Leeflang MM, Takwoingi Y (editors). Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy. Version 2.0 (updated July 2023). Cochrane, 2023. Available from training.cochrane.org/handbook-diagnostic-test-accuracy/current.

Dufour 2000

Dufour DR, Lott JA, Nolte FS, Gretch DR, Koff RS, Seeff LB. Diagnosis and monitoring of hepatic injury. I. Performance characteristics of laboratory tests. *Clinical Chemistry* 2000;**46**(12):2027-49.

EASL 2015

European Association for the Study of the Liver. EASL recommendations on treatment of hepatitis C 2015. Journal of Hepatology 2015;**63**(1):199-236. [DOI: 10.1016/j.jhep.2015.03.025]

EASL 2020

Pawlotsky JM, Negro F, Aghemo A, Berenguer M, Dalgard O, Dusheiko G, et al, European Association for the Study of the Liver. EASL recommendations on treatment of hepatitis C: final update of the series. *Journal of Hepatology* 1;**73**(5):1170-218. [DOI: https://doi.org/10.1016/j.jhep.2020.08.018]

EASL 2021

Berzigotti A, Tsochatzis E, Boursier J, Castera L, Cazzagon N, Friedrich-Rust M, et al, European Association for the Study of the Liver. EASL Clinical Practice Guidelines on non-invasive tests for evaluation of liver disease severity and prognosis – 2021 update. *Journal of Hepatology* 1;**75**(3):659-89. [DOI: 10.1016/j.jhep.2021.05.025]

EASL-ALEH 2015

European Association for the Study of the Liver. EASL-ALEH clinical practice guidelines: non-invasive tests for evaluation of liver disease severity and prognosis. Journal of Hepatology 2015;**63**(1):237-64. [DOI: 10.1016/j.jhep.2015.04.006]

Francque 2021

Francque SM, Bedossa P, Ratziu V, Anstee QM, Bugianesi E, Sanyal AJ, et al. A randomized, controlled trial of the pan-PPAR agonist lanifibranor in NASH. *New England Journal of Medicine* 2021;**385**(17):1547-58. [DOI: 10.1056/NEJMoa2036205]

Glanville 2010

Glanville J, Cikalo M, Crawford F, Dozier M, Lowson P. Handsearching for reports of diagnostic test accuracy studies: adding to the evidence base. In: Bero L, Montgomery P, Robinson K, Pigott T, Krause K. Bringing Evidence-Based Decision-Making to New Heights. Abstracts of the 2010 Joint Colloquium of The Cochrane and Campbell Collaborations; 2010 Oct 18-22; Keystone, USA. John Wiley & Sons; 2010. [DOI: 10.1002/14651858.CD000002]

Goodman 2007

Goodman ZD. Grading and staging systems for inflammation and fibrosis in chronic liver diseases. *Journal of Hepatology* 1;**47**(4):598-607.

GRADEpro GDT [Computer program]

GRADEpro GDT. Version accessed 23 April 2024. Hamilton (ON): McMaster University (developed by Evidence Prime), 2010. Available at gradepro.org.

Greenhalgh 2005

Greenhalgh T, Peacock R. Effectiveness and efficiency of search methods in systematic reviews of complex evidence: audit of primary sources. *BMJ (Clinical Research Ed.)* 2005;**331**(7524):1064-5.

Grigorescu 2006

Grigorescu M. Noninvasive biochemical markers of liver fibrosis. *Journal of Gastrointestinal and Liver Diseases* 2006;**15**(2):149-59.

Guyatt 2011

Guyatt GH, Oxman AD, Schünemann HJ, Tugwell P, Knottnerus A. GRADE guidelines: a new series of articles in the Journal of Clinical Epidemiology. *Journal of Clinical Epidemiology* 2011;**64**(4):380-2. [DOI: 10.1016/j.jclinepi.2010.09.011]

Hanley 1982

Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 1982;**143**(1):29-36. [DOI: 10.1148/radiology.143.1.7063747]

Huwart 2008

Huwart L, Sempoux C, Vicaut E, Salameh N, Annet L, Danse E, et al. Magnetic resonance elastography for the noninvasive staging of liver fibrosis. *Gastroenterology* 2008;**135**(1):32-40.

Ishak 1995

Ishak K, Baptista A, Bianchi L, Callea F, De Groote J, Gudat F, et al. Histological grading and staging of chronic hepatitis. *Journal of Hepatology* 1995;**22**(6):696-9.

Ito 2023

Ito T, Nguyen VH, Tanaka T, Park H, Yeh ML, Kawanaka M, et al. Poor diagnostic efficacy of noninvasive tests for advanced fibrosis in obese or younger than 60 diabetic NAFLD patients. *Clinical Gastroenterology and Hepatology* 2023;**21**(4):1013-22. [DOI: 10.1016/j.cgh.2022.05.015]

Kalambokis 2007

Kalambokis G, Manousou P, Vibhakorn S, Marelli L, Cholongitas E, Senzolo M, et al. Transjugular liver biopsy - indications, adequacy, quality of specimens, and complications - a systematic review. *Journal of Hepatology* 2007;**47**(2):284-94.

Knodell 1981

Knodell RG, Ishak KG, Black WC, Chen TS, Craig R, Kaplowitz N, et al. Formulation and application of a numerical scoring system for assessing histological activity in asymptomatic chronic active hepatitis. *Hepatology* 1981;**1**(5):431-5.

Li 2017

Li Q, Lu C, Li W, Huang Y, Chen L. Impact of age on the diagnostic performances and cut-offs of APRI and FIB-4 for significant fibrosis and cirrhosis in chronic hepatitis B. *Oncotarget* 2017;**8**(28):45768. [DOI: 10.18632/oncotarget.17470]

Lichtinghagen 2013

Lichtinghagen R, Pietsch D, Bantel H, Manns MP, Brand K, Bahr MJ. The Enhanced Liver Fibrosis (ELF) score: normal values, influence factors and proposed cut-off values. *Journal of Hepatology* 2013;**59**(2):236-42.

Mathes 2019

Mathes T, Pieper D. An algorithm for the classification of study designs to assess diagnostic, prognostic and predictive test accuracy in systematic reviews. *Systematic Reviews* 2019;**8**:226. [DOI: 10.1186/s13643-019-1131-4]

McPherson 2017

McPherson S, Hardy T, Dufour JF, Petta S, Romero-Gomez M, Allison M, et al. Age as a confounding factor for the accurate non-invasive diagnosis of advanced NAFLD fibrosis. *American Journal of Gastroenterology* 2017;**112**(5):740. [DOI: 10.1038/ajg.2016.453]

METAVIR 1996

Bedossa P, Poynard T. An algorithm for the grading of activity in chronic hepatitis C. The METAVIR Cooperative Study Group. *Hepatology* 1996;**24**(2):289-93.

Nguyen 2011

Nguyen D, Talwalkar JA. Noninvasive assessment of liver fibrosis. *Hepatology* 2011;**53**(6):2107-10.

Pavlov 2015a

Pavlov CS, Casazza G, Nikolova D, Tsochatzis E, Ivashkin VT, Gluud C. FibroTest, transient elastography method, and combined FibroTest and transient elastography method for diagnosis of severe hepatic fibrosis and cirrhosis in adults with chronic hepatitis C. *Cochrane Database of Systematic Reviews* 2016, Issue 7. Art. No: CD012291. [DOI: 10.1002/14651858.CD012291]

Pavlov 2015b

Pavlov CS, Casazza G, Nikolova D, Tsochatzis E, Burroughs AK, Ivashkin VT, et al. Transient elastography for diagnosis of stages of hepatic fibrosis and cirrhosis in people with alcoholic liver disease. *Cochrane Database of Systematic Reviews* 2015, Issue 1. Art. No: CD010542. [DOI: 10.1002/14651858.CD010542.pub2]

Poynard 2011

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

Saadoun 2006

Saadoun D, Asselah T, Resche-Rigon M, Charlotte F, Bedossa P, Valla D, et al. Cryoglobulinemia is associated with steatosis and fibrosis in chronic hepatitis C. *Hepatology* 2006;**43**(6):1337-45. [DOI: 10.1002/hep.21190]

Scheuer 1991

Scheuer PJ. Classification of chronic viral hepatitis: a need for reassessment. *Journal of Hepatology* 1991;**13**(3):372-4.

Sebastiani 2006

Sebastiani G, Alberti A. Non invasive fibrosis biomarkers reduce but not substitute the need for liver biopsy. *World Journal of Gastroenterology* 2006;**12**(23):3682-94.

Shah 2009

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. *Clinical Gastroenterology and Hepatology* 2009;**7**(10):1104-12.

Shaheen 2008

Shaheen AA, Myers RP. Systematic review and meta–analysis of the diagnostic accuracy of fibrosis marker panels in patients with HIV/hepatitis C coinfection. *HIV Clinical Trials* 2008;**9**(1):43-51. [DOI: 10.1310/hct0901-43]

Smith 2009

Smith JO, Sterling RK. Systematic review: non-invasive methods of fibrosis analysis in chronic hepatitis C. *Alimentary Pharmacology & Therapeutics* 2009;**30**(6):557-76.

Standish 2006

Standish RA, Cholongitas E, Dhillon A, Burroughs AK, Dhillon AP. An appraisal of the histopathological assessment of liver fibrosis. *Gut* 2006;**55**(4):569-78.

Tsochatzis 2014

Tsochatzis EA, Bosch J, Burroughs AK. Liver cirrhosis. *Lancet* 2014;**383**(9930):1749-61.

Whiting 2011

Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Annals of Internal Medicine* 2011;**155**(8):529-36.

Xiao 2015

Xiao G, Yang J, Yan L. Comparison of diagnostic accuracy of aspartate aminotransferase to platelet ratio index and fibrosis-4 index for detecting liver fibrosis in adult patients with chronic hepatitis B virus infection: a systemic review and meta-analysis. *Hepatology* 2015;**61**(1):292-302.

References to other published versions of this review

Kalafateli 2015

Kalafateli M, Gurusamy KS, Noel-Storr AH, Burroughs AK, Tsochatzis E. Forns index and 'FIB4' for staging of fibrosis in adults with chronic hepatitis C. *Cochrane Database of Systematic Reviews* 2015, Issue 11. Art. No: CD011929. [DOI: 10.1002/14651858.CD011929]

CHARACTERISTICS OF STUDIES

Characteristics of included studies *[ordered by study ID]*

Abdel-Hameed 2021a

Study characteristics	
Patient Sampling	Study design: cross-sectional diagnostic test accuracy study Sampling method: cohort-based Direction of data collection: retrospective Country: USA Inclusion criteria: infected with HIV and/or HCV and had a biopsy report with serum samples collected at the time of the biopsy Exclusion criteria: biopsy record showed other liver pathology or if the biopsy was inadequate
Patient characteristics and setting	Centre details: multicentre, University of Cincinnati and University of Maryland Sample size: 245 Mean age: 54 Gender (% male): 60 (HCV cohort) and 79 (HCV/HIV cohort) Mean BMI: not reported Mean ALT: 53 Special characteristics: all participants were co-infected with HIV
Index tests	Test name(s): FIB-4 Threshold(s) used: HCV cohort F1+ 1.33, F3+ 1.51, F4 2.63; HCV+HIV cohort F1+ 1.14, F3+ 1.6, F4 1.84
Target condition and reference standard(s)	Target condition(s): F1+, F3+, F4 Reference standard: liver biopsy Quality of liver biopsy: at least 11 portal areas and not fragmented
Flow and timing	Flow: all participants received the index test and none were excluded from the analysis Time between index test and biopsy: same day
Comparative	Comparators: ELF index and APRI score
Notes	We included two cohorts (HCV only and HCV + HIV) in the review and represent them under 'a' and 'b', respectively. We extracted separate data on these cohorts for the purposes of our meta-analysis.

Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Unclear risk	

Abdel-Hameed 2021a *(Continued)*

Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index test (FIB-4)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index test (Forns)			
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Low risk	

Abdel-Hameed 2021b

Study characteristics	
Patient Sampling	Study design: cross-sectional diagnostic test accuracy study Sampling method: cohort-based Direction of data collection: retrospective Country: USA Inclusion criteria: infected with HIV and/or HCV and had a biopsy report with serum samples collected at the time of the biopsy Exclusion criteria: biopsy record showed other liver pathology or if the biopsy was inadequate

Abdel-Hameed 2021b *(Continued)*

Patient characteristics and setting	Centre details: Multicentre, University of Cincinnati and University of Maryland Sample size: 245 Mean age: 54 Gender (% male): 60 (HCV cohort) and 79 (HCV/HIV cohort) Mean BMI: Not reported Mean ALT: 53 Special characteristics: all patients were co-infected with HIV
Index tests	Test name(s): FIB-4 Threshold(s) used: HCV cohort F1+ 1.33, F3+ 1.51, F4 2.63; HCV+HIV cohort F1+ 1.14, F3+ 1.6, F4 1.84
Target condition and reference standard(s)	Target condition(s): F1+, F3+, F4 Reference standard: liver biopsy Quality of liver biopsy: at least 11 portal areas and not fragmented
Flow and timing	Flow: all patients received the index test and none were excluded from the analysis Time between index test and biopsy: same day
Comparative	Comparators: ELF index and APRI score
Notes	We included two cohorts (HCV only and HCV + HIV) in the review and represent them under 'a' and 'b', respectively. We extracted separate data on these cohorts for the purposes of our meta-analysis.

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Unclear risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index test (FIB-4)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	

Abdel-Hameed 2021b *(Continued)*

Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index test (Forns)			
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Low risk	

Abdelsameea 2020

Study characteristics	
Patient Sampling	Study design: cross-sectional diagnostic test accuracy study Sampling method: cohort-based Direction of data collection: not reported Country: Egypt Inclusion criteria: people with CHC who were treatment eligible (for interferon or direct-acting antivirals) and undergoing liver biopsy in preparation for treatment Exclusion criteria: people co-infected with other viruses, having other liver disease, decompensated liver cirrhosis, pregnancy, hepatocellular carcinoma or being treatment ineligible
Patient characteristics and setting	Centre details: single centre, National Liver Institute hospitals, Menoufia University Sample size: 1310 Mean age: 37.5 Gender (% male): 71 Mean BMI: not reported Mean ALT: 48
Index tests	Test name(s): FIB-4 Threshold(s) used: F1+ 0.65, F2+ 1.28, F3+ 1.3, F4 1.3

Abdelsameea 2020 *(Continued)*

Target condition and reference standard(s)	Target condition(s): F1+, F2+, F3+, F4 Reference standard: liver biopsy Quality of liver biopsy: at least 16 mm to 22mm long and containing at least six portal tracts
Flow and timing	Flow: all participants received the index test and none were excluded from the analysis Time between index test and biopsy: same day or day preceding the biopsy
Comparative	Comparators: APRI and GUCI scores
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Unclear risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index test (FIB-4)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index test (Forns)			
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	

Abdelsameea 2020 *(Continued)*

Are there concerns that the target condition as defined by the reference standard does not match the question?		Low concern
DOMAIN 4: Flow and timing		
Was there an appropriate interval between index test and reference standard?	Yes	
Did all patients receive the same reference standard?	Yes	
Were all patients included in the analysis?	Yes	
Could the patient flow have introduced bias?		Low risk

Abo El-Khair 2019

Study characteristics	
Patient Sampling	Study design: cross-sectional diagnostic test accuracy study Sampling method: cohort-based Direction of data collection: not reported Country: Egypt Inclusion criteria: HCV pre-treatment patients with liver biopsy in accordance to APASL recommendation Exclusion criteria: < 18 years, HBV and/or HIV infection, AIH, decompensated liver cirrhosis, HCC, people with diabetes, anaemia, renal disease, or haemolytic disorders, people treated with erythropoietin, prednisolone or hepatotoxic drugs
Patient characteristics and setting	Centre details: single centre, Tropical Medicine Department at Mansoura University Sample size: 100 Mean age: 45 Gender (% male): 62 Mean BMI: not reported Mean ALT: not reported
Index tests	Test name(s): FIB-4 Threshold(s) used: F2+ 1.17, F3+ 1.84, F4 1.60
Target condition and reference standard(s)	Target condition(s): F2+, F3+, F4 Reference standard: liver biopsy Quality of liver biopsy: not reported
Flow and timing	Flow: all participants received the index test and none were excluded from the analysis Time between index test and biopsy: same day
Comparative	Comparators: novel Fibrosis Prediction Score (FPS) and APRI
Notes	
Methodological quality	

Abo El-Khair 2019 *(Continued)*

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Unclear risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index test (FIB-4)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index test (Forns)			
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Low risk	

Ahmad 2011

Study characteristics	
Patient Sampling	Study design: cross-sectional diagnostic test accuracy study Sampling method: cohort-based Direction of data collection: retrospective Country: Pakistan Inclusion criteria: HCV RNA-positive patients were identified among HCV antibody (anti- HCV) positive patients Exclusion criteria: previous interferon therapy or immunosuppressants, HBV, HIV, liver cancer, refusal of or contraindication to liver biopsy
Patient characteristics and setting	Centre details: multicentre, Jinnah Hospital, Lahore, Mayo Hospital, Lahore and Liver Centre, Faisalabad Sample size: 157 Mean age: 38.1 Gender (% male): 73 Mean BMI: not reported Mean ALT: 134
Index tests	Test name(s): FIB-4 Threshold(s) used: F3+ 1.45, F3+ 3.25
Target condition and reference standard(s)	Target condition(s): F3+ Reference standard: liver biopsy Quality of liver biopsy: unclear
Flow and timing	Flow: all participants received the index test and none were excluded from the analysis Time between index test and biopsy: not reported
Comparative	Comparators: AAR, APRI, Forns index
Notes	

Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Unclear risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index test (FIB-4)			

Ahmad 2011 *(Continued)*

Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index test (Forns)			
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Unclear risk	

Alboraie 2015

Study characteristics	
Patient Sampling	Study design: cross-sectional diagnostic test accuracy study Sampling method: cohort-based Direction of data collection: retrospective Country: Egypt Inclusion criteria: interferon naïve Exclusion criteria: chronic HBV co-infection
Patient characteristics and setting	Centre details: single centre Kasr Al-Aini Viral Hepatitis Center, University of Cairo Sample size: 100 Mean age: 40 Gender (% male): 67 Mean BMI: not reported

Alboraie 2015 *(Continued)*

	Mean ALT: 55
Index tests	Test name(s): FIB-4 and Forns index Threshold(s) used: FIB-4 F2+ 1, F3+ 1.45, F4 2.25. Forns index F2+ 4.02.
Target condition and reference standard(s)	Target condition(s): F2+, F3+, F4 Reference standard: liver biopsy Quality of liver biopsy: at least four portal tracts
Flow and timing	Flow: all participants received the index test and none were excluded from the analysis Time between index test and biopsy: blood samples were taken at the same time as the liver biopsy
Comparative	Comparators: Egy-score, APRI
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Unclear risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index test (FIB-4)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index test (Forns)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		

Alboraie 2015 *(Continued)*

Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Low risk	

Amorim 2012

Study characteristics	
Patient Sampling	Study design: cross-sectional diagnostic test accuracy study Sampling method: cohort-based Direction of data collection: retrospective Country: Brazil Inclusion criteria: HCV infection with complete biopsy and biochemical data Exclusion criteria: HBV infection, HIV co-infection, NAFLD, autoimmune hepatitis, insufficient liver biopsy
Patient characteristics and setting	Centre details: single centre, University of Catarina, Brazil Sample size: 119 Mean age: 43.7 Gender (% male): 62 Mean BMI: not reported Mean ALT: 78
Index tests	Test name(s): FIB-4 Threshold(s) used: F2+ 1.45, 3.25
Target condition and reference standard(s)	Target condition(s): F2+ Reference standard: liver biopsy

Amorim 2012 *(Continued)*

	Quality of liver biopsy: not reported
Flow and timing	Flow: all participants meeting inclusion criteria were included in the analysis Time between index test and biopsy: within 6 months
Comparative	Comparators: APRI score and AST/ALT ratio
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Low risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index test (FIB-4)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index test (Forns)			
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern

Amorim 2012 *(Continued)*

DOMAIN 4: Flow and timing	
Was there an appropriate interval between index test and reference standard?	No
Did all patients receive the same reference standard?	Yes
Were all patients included in the analysis?	Yes
Could the patient flow have introduced bias?	High risk

Andrés-Otero 2016

Study characteristics	
Patient Sampling	Study design: cross-sectional diagnostic test accuracy study Sampling method: cohort-based Direction of data collection: prospective Country: Spain Inclusion criteria: CHC established by the presence of HCV RNA using polymerase chain reaction assays Exclusion criteria: alcohol excess, renal insufficiency, other causes of liver disease, decompensated cirrhosis, HCC, transplant recipient, prior interferon therapy, poor biopsy quality
Patient characteristics and setting	Centre details: single centre, University Hospital “Lozano Blesa” of Zaragoza (Spain) Sample size: 84 Mean age: 44 Gender (% male): 82 Mean BMI: not reported Mean ALT: not reported
Index tests	Test name(s): FIB-4 and Forns index Threshold(s) used: FIB-4 F2+ 1.5, F3+ 2.05, F4 2.63; Forns index F2+ 5.4, F3+ 6.66, F4 6.66
Target condition and reference standard(s)	Target condition(s): F2+, F3+, F4 Reference standard: liver biopsy Quality of liver biopsy: 15 mm length (number of portal tracts not reported)
Flow and timing	Flow: all participants received the index test and none were excluded from the analysis Time between index test and biopsy: same day
Comparative	Comparators: AAR, APRI, Fibroindex, MODEL3 score, GUCI, Forns index, fibrosis cirrhosis index (FCI), Pohl score, AP index (age and platelet), cirrhosis discriminant score (CDS), Hospital Gregorio Marañón scores (HGM-1 and HGM-2)
Notes	
Methodological quality	

Andrés-Otero 2016 *(Continued)*

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Low risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index test (FIB-4)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index test (Forns)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and timing			

Andrés-Otero 2016 *(Continued)*

Was there an appropriate interval between index test and reference standard?	Yes	
Did all patients receive the same reference standard?	Yes	
Were all patients included in the analysis?	Yes	
Could the patient flow have introduced bias?		Low risk

Attallah 2012

Study characteristics	
Patient Sampling	Study design: cross-sectional diagnostic test accuracy study Sampling method: cohort-based Direction of data collection: prospective Country: Egypt Inclusion criteria: HCV-RNA using polymerase chain reaction (PCR). Cirrhotic patients were compensated at the time of inclusion. Exclusion criteria: HBV, prior antiviral or immunosuppressive therapy, decompensation, typhoid, B12 deficiency, leukaemia
Patient characteristics and setting	Centre details: single centre, Mansoura University hospitals, Mansoura, Egypt Sample size: 3212 Mean age: 42.5 Gender (% male): 76 Mean BMI: not reported Mean ALT: 52
Index tests	Test name(s): FIB-4 Threshold(s) used: F3+ 3.25
Target condition and reference standard(s)	Target condition(s): F3+ Reference standard: liver biopsy Quality of liver biopsy: at least 15 mm and/or contains at least five portal tracts, except for cirrhosis, for which no limitation was required
Flow and timing	Flow: all participants received the index test and none were excluded from the analysis Time between index test and biopsy: within 2 weeks
Comparative	Comparators: novel score Fibrosis Routine Test (FRT), APRI, Lok index, GUCI, Forns index, FibroQ, Fibrosis cirrhosis index (FCI), four routine laboratory blood tests (4RLB)
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns

Attallah 2012 *(Continued)*

DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Unclear risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index test (FIB-4)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index test (Forns)			
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Low risk	

Baldwin 2020

Study characteristics	
Patient Sampling	Study design: cross-sectional diagnostic test accuracy study Sampling method: cohort-based Direction of data collection: retrospective Country: USA Inclusion criteria: consecutive participants with biopsies for HCV Exclusion criteria: biopsies of liver masses, transplanted organs, severe necrosis, multiple aetiologies
Patient characteristics and setting	Centre details: single centre, University of Alabama Sample size: 518 Mean age: not reported Gender (% male): not reported Mean BMI: not reported Mean ALT: not reported
Index tests	Test name(s): FIB-4 Threshold(s) used: F3+ 1.45, 3.25
Target condition and reference standard(s)	Target condition(s): F3+ Reference standard: liver biopsy Quality of liver biopsy: not reported
Flow and timing	Flow: 199 participants had both index test and reference standard but were not reported in the analysis Time between index test and biopsy: within 6 months
Comparative	Comparators: NAFLD score, APRI score
Notes	

Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Low risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index test (FIB-4)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		

Baldwin 2020 *(Continued)*

Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index test (Forns)			
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and timing			
Was there an appropriate interval between index test and reference standard?	No		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	No		
Could the patient flow have introduced bias?		High risk	

Bonnard 2015

Study characteristics	
Patient Sampling	Study design: cross-sectional diagnostic test accuracy study Sampling method: cohort-based Direction of data collection: prospective Country: Egypt Inclusion criteria: adult with HCV infection Exclusion criteria: pregnant women, people infected with a non-4 genotype, positive hepatitis B surface (HBs) antigen, contra-indication to biopsy, decompensation
Patient characteristics and setting	Centre details: single centre, National Hepatology and Tropical Medicine Research Institute (Cairo, Egypt) Sample size: 500 Mean age: 39.4 Gender (% male): 64 Mean BMI: 27.9 Mean ALT: 1.61 x upper limit of normal Special characteristics: all participants were infected with HCV genotype 4, with other genotypes excluded

Bonnard 2015 *(Continued)*

Index tests	Test name(s): FIB-4 Threshold(s) used: F2+ 1, F4 1.27
Target condition and reference standard(s)	Target condition(s): F2+, F4 Reference standard: liver biopsy Quality of liver biopsy: ≥ 15 mm length or ≥ 10 portal tracts
Flow and timing	Flow: 162 participants had both liver biopsy and index test and were not included in any analysis Time between index test and biopsy: within 15 days
Comparative	Comparators: APRI, FibroTest, FibroScan
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Unclear risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index test (FIB-4)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index test (Forns)			
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		

Bonnard 2015 *(Continued)*

Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk
Are there concerns that the target condition as defined by the reference standard does not match the question?		Low concern
DOMAIN 4: Flow and timing		
Was there an appropriate interval between index test and reference standard?	Yes	
Did all patients receive the same reference standard?	Yes	
Were all patients included in the analysis?	No	
Could the patient flow have introduced bias?		High risk

Bourliere 2006

Study characteristics	
Patient Sampling	Study design: cross-sectional diagnostic test accuracy study Sampling method: cohort-based Direction of data collection: prospective Country: France Inclusion criteria: HCV infection per serology Exclusion criteria: cholesterol test not available
Patient characteristics and setting	Centre details: multicentre, Saint-Joseph Hospital and La Conception Hospital (Marseille), Archet Hospital (Nice), Hyères Hospital (Hyères) and Arnault Tzanck Institute (St. Laurent du Var) Sample size: 520 Mean age: 46 Gender (% male): not reported Mean BMI: note reported Mean ALT: 75.2
Index tests	Test name(s): Forns index Threshold(s) used: 4.21, 6.9
Target condition and reference standard(s)	Target condition(s): F2+ Reference standard: liver biopsy Quality of liver biopsy: 15 mm or more in length, five or more portal tracts and one fragment
Flow and timing	Flow: 235 were included in the analysis from an initial cohort of 520 due to unavailability of serum tests Time between index test and biopsy: same day
Comparative	Comparators: FibroTest, APRI
Notes	
Methodological quality	

Bourliere 2006 *(Continued)*

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Low risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index test (FIB-4)			
DOMAIN 2: Index test (Forns)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	No		
Could the patient flow have introduced bias?		High risk	

Calès 2010

Study characteristics	
Patient Sampling	Study design: cross-sectional diagnostic test accuracy study Sampling method: cohort-based Direction of data collection: prospective Country: France Inclusion criteria: anti-HCV and anti-HIV antibodies, and HCV RNA in serum Exclusion criteria: additional causes of liver disease, particularly HBV co-infection, complicated cirrhosis, anti-fibrotic treatment in the previous 6 months, excess alcohol consumption
Patient characteristics and setting	Centre details: multicentre, four tertiary centres, Angers, Paris Hôpital Européen Georges Pompidou, Rennes and Tours, and one secondary centre, La Roche sur Yon Sample size: 467 Mean age: 41.4 Gender (% male): 64 Mean BMI: not reported Mean ALT: not reported Special characteristics: all participants were co-infected with HIV
Index tests	Test name(s): FIB-4 Threshold(s) used: 1.28
Target condition and reference standard(s)	Target condition(s): F2+ Reference standard: liver biopsy Quality of liver biopsy: not reported
Flow and timing	Flow: all participants were included in the analysis Time between index test and biopsy: immediately before or no more than 3 months after the liver biopsy was performed
Comparative	Comparators: APRI, FibroTest, Hepascore, FibroMeter, human immunodeficiency and C virus (HICV) test, FibroMeter HICV
Notes	

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Low risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern

Calès 2010 *(Continued)*

DOMAIN 2: Index test (FIB-4)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index test (Forns)			
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Low risk	

Cheng 2019

Study characteristics	
Patient Sampling	Study design: cross-sectional diagnostic test accuracy study Sampling method: cohort-based Direction of data collection: retrospective Country: Taiwan Inclusion criteria: persistent viraemia for at least six months Exclusion criteria: HIV, Wilson's disease, primary biliary cirrhosis, haemochromatosis, and autoimmune hepatitis, hematological diseases, liver transplantation for any indication, decompensated cirrhosis or hepatic failure
Patient characteristics and setting	Centre details: single centre, Tatung Mackay Memorial Hospital

Cheng 2019 *(Continued)*

	Sample size: 113 Mean age: 60.2 Gender (% male): 55 Mean BMI: 26.5 Mean ALT: 70.7
Index tests	Test name(s): FIB-4 Threshold(s) used: F3+ 2.6, F4 4.018
Target condition and reference standard(s)	Target condition(s): F3+, F4 Reference standard: liver biopsy Quality of liver biopsy: minimum of 1.0 cm in size and a minimum of four portal tracts
Flow and timing	Flow: all participants received the index test and none were excluded from the analysis Time between index test and biopsy: same day
Comparative	Comparators: APRI score
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Unclear risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index test (FIB-4)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index test (Forns)			
DOMAIN 3: Reference standard			

Cheng 2019 *(Continued)*

Is the reference standard likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Low risk	

Conti 2019

Study characteristics	
Patient Sampling	Study design: cross-sectional diagnostic test accuracy study Sampling method: cohort-based Direction of data collection: retrospective Country: Italy Inclusion criteria: consecutive participants with chronic liver disease of any cause scheduled for liver biopsy (subset of HCV-only participants available in analysis) Exclusion criteria: younger than 18 years old, previous liver transplant, decompensated cirrhosis, HCC, acute liver injury, time between liver stiffness measurement (LSM) and biopsy exceeded 2 weeks, biopsy sample smaller than 20 mm or < 11 portal tracts on biopsy
Patient characteristics and setting	Centre details: single centre, Diagnostic and Interventional Ultrasound Unit 144 of Policlinico S. Orsola-Malpighi, Bologna, Italy Sample size: 491 Mean age: 52 Gender (% male): 90 Mean BMI: 24.7 Mean ALT: 53
Index tests	Target condition(s): F3+ Reference standard: liver biopsy Quality of liver biopsy: > 20 mm length or > 10 portal tracts on biopsy
Target condition and reference standard(s)	Flow: 361 included from a total of 491, with the rest excluded due to uninterpretable or indeterminate results Time between index test and biopsy: within 2 weeks

Conti 2019 *(Continued)*

Flow and timing	Flow: 361 included from a total of 491, with the rest excluded due to uninterpretable or indeterminate results Time between index test and biopsy: within 2 weeks
Comparative	Comparators: ElastPQ, TE, APRI
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Low risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index test (FIB-4)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index test (Forns)			
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and timing			

Conti 2019 *(Continued)*

Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	No		
Could the patient flow have introduced bias?		High risk	

Cordie 2018

Study characteristics			
Patient Sampling	Study design: cross-sectional diagnostic test accuracy study Sampling method: cohort-based Direction of data collection: not reported Country: Egypt Inclusion criteria: chronic HCV patients naive to therapy; liver biopsy of at least 15 mm and 4 portal tracts Exclusion criteria: < 18 years, decompensated cirrhosis, HCC or other liver diseases, people treated for HCV		
Patient characteristics and setting	Centre details: two centres, Kasr Al-Aini Viral Hepatitis Center, Faculty of Medicine, Cairo University and National Hepatology & Tropical Medicine Research Institute Sample size: 200 Mean age: 40 Gender (% male): 65 Mean BMI: 26.8 Mean ALT: 54.2		
Index tests	Test name(s): FIB-4 Threshold(s) used: 1.27		
Target condition and reference standard(s)	Target condition(s): F2+ Reference standard: liver biopsy Quality of liver biopsy: 15 mm in length with minimum of 4 portal tracts		
Flow and timing	Flow: all participants received the index test and none were excluded from the analysis Time between index test and biopsy: same day		
Comparative	Comparators: Fib-4, Egy-score, Aspartate-to-platelet ratio index (APRI), and Göteborg University Cirrhosis Index (GUCI)		
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient selection			

Cordie 2018 *(Continued)*

Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Low risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index test (FIB-4)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index test (Forns)			
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Low risk	

Corradi 2009

Study characteristics			

Corradi 2009 *(Continued)*

Patient Sampling	Study design: case-control diagnostic test accuracy study Sampling method: case-control Direction of data collection: prospective Country: Italy Inclusion criteria: confirmation of recurrent HCV hepatitis in transplant recipient Exclusion criteria: co-infection with HIV, HBV, autoimmune hepatitis, other causes of liver disease, obesity (BMI > 35), alcohol excess, ascites
Patient characteristics and setting	Centre details: single centre, Division of Internal Medicine, University of Bologna Sample size: 56 Mean age: 57.5 Gender (% male): 83 Mean BMI: 24.5 Mean ALT: 72 Special characteristics: transplant recipients
Index tests	Test name(s): Forns index Threshold(s) used: F2+ 4.2 and 9
Target condition and reference standard(s)	Target condition(s): F2+ Reference standard: liver biopsy Quality of liver biopsy: minimum length of biopsy 16mm with 9 portal tracts
Flow and timing	Flow: 20 of 56 participants had incomplete serological parameters and were not included in full analysis Time between index test and biopsy: within 4 weeks
Comparative	Comparators: liver stiffness, FibroTest, APRI, Forns index, Benlloch index
Notes	

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index test (FIB-4)			
DOMAIN 2: Index test (Forns)			

Corradi 2009 *(Continued)*

Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	No		
Could the patient flow have introduced bias?		High risk	

Crisan 2012

Study characteristics	
Patient Sampling	Study design: cross-sectional diagnostic test accuracy study Sampling method: cohort-based Direction of data collection: prospective Country: Romania Inclusion criteria: diagnosed with CHC and underwent liver biopsy Exclusion criteria: hepatitis B, autoimmune liver disease, Wilson disease, haemochromatosis, α1-antitripsin deficiency, HIV infection; history of hepatotoxic or steatosis-inducing drug use, alcohol excess
Patient characteristics and setting	Centre details: single centre, 3rd Medical Clinic, Cluj- Napoca, Romania Sample size: 446 Mean age: 49 Gender (% male): 38 Mean BMI: 27

Crisan 2012 *(Continued)*

	Mean ALT: 76
Index tests	Test name(s): FIB-4 and Forns index Threshold(s) used: FIB-4 F2+ 1.26, F3+ 3.74; Forns index F2+ 4.47, F3+ 7.3
Target condition and reference standard(s)	Target condition(s): F2+, F3+ Reference standard: liver biopsy Quality of liver biopsy: median length 11 mm but number of portal tracts not reported
Flow and timing	Flow: all participants received the index test and none were excluded from the analysis Time between index test and biopsy: not reported
Comparative	Comparators: APRI, FibroTest, Hepascore, and FibroMeter scores and TE (FibroScan)
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Low risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index test (FIB-4)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index test (Forns)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		

Crisan 2012 *(Continued)*

Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Unclear risk	

De Oliveira 2016

Study characteristics	
Patient Sampling	Study design: cross-sectional diagnostic test accuracy study Sampling method: cohort-based Direction of data collection: retrospective Country: Brazil Inclusion criteria: hepatitis C virus infection confirmed by PCR Exclusion criteria: HBV, HIV, liver disease of other aetiology, HCC, previous antiviral therapy or transplant, immunosuppressants, alcohol excess, insufficient liver biopsy. People with clinical, endoscopic, or sonographical evidence of cirrhosis were excluded.
Patient characteristics and setting	Centre details: single centre, Department of Internal Medicine, Federal University of São Paulo Sample size: 798 Mean age: 56.9 Gender (% male): 50.4 Mean BMI: not reported Mean ALT: 86.9
Index tests	Test name(s): FIB-4 Threshold(s) used: 1.45, 3.27

De Oliveira 2016 *(Continued)*

Target condition and reference standard(s)	Target condition(s): F2+, F3+, F4 Reference standard: liver biopsy Quality of liver biopsy: not reported
Flow and timing	Flow: all participants received the index test and none were excluded from the analysis Time between index test and biopsy: the mean interval was less than 3 months (but by inference not all participants were less than 3 months)
Comparative	Comparators: APRI
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index test (FIB-4)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index test (Forns)			
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	

De Oliveira 2016 *(Continued)*

Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and timing			
Was there an appropriate interval between index test and reference standard?	No		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Demma 2018

Study characteristics	
Patient Sampling	Study design: cross-sectional diagnostic test accuracy study Sampling method: cohort-based Direction of data collection: retrospective and prospective Country: UK Inclusion criteria: HCV Exclusion criteria: not reported (abstract only)
Patient characteristics and setting	Centre details: single centre, Royal Free Hospital London Sample size: 612 Mean age: not reported Gender (% male): not reported Mean BMI: not reported Mean ALT: not reported
Index tests	Test name(s): FIB-4 Threshold(s) used: F3+ 1.45, 3.25; F4 1.45, 3.25
Target condition and reference standard(s)	Target condition(s): F3+, F4 Reference standard: 100 received biopsy, 512 received LSM (analysed separately in unpublished data sought from authors) Quality of liver biopsy: not reported
Flow and timing	Flow: all participants received the index test and none were excluded from the analysis. 512 participants did not receive biopsy. Time between index test and biopsy: unclear; states "contemperaneous"
Comparative	Comparators: Elift, APRI
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient selection			

Demma 2018 *(Continued)*

Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		Unclear risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index test (FIB-4)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index test (Forns)			
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	No		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Ferenci 2014

Study characteristics

Ferenci 2014 *(Continued)*

Patient Sampling	Study design: cross-sectional diagnostic test accuracy study Sampling method: cohort-based Direction of data collection: prospective Countries: 19 countries including Austria, Belgium, USA, Brazil, Ireland, Canada, Italy, Croatia, United Kingdom, France, Hungary, Macedonia, Mexico, Morocco, Poland, Romania, Serbia, Slovenia, Sweden Inclusion criteria: HCV treatment-naive patients with genotype 1 HCV mono-infection prescribed only peginterferon alfa-2a/ribavirin and participating in PROPHESYS 1–3 trials Exclusion criteria: HBV and HIV co-infection, previous IFN treatment, autoimmune hepatitis, decompensated liver disease, unstable or uncontrolled cardiac disease
Patient characteristics and setting	Centre details: multicentre in 19 countries, mix of centre types Sample size: 1592 Mean age: 48 Gender (% male): 813 Mean BMI: 26.2 Mean ALT: 69.9 Special characteristics: all participants were genotype 1
Index tests	Test name(s): FIB-4 Threshold(s) used: 0.68, 1.4, 3.2
Target condition and reference standard(s)	Target condition(s): F1+, F2+, F3+, F4 Reference standard: liver biopsy Quality of liver biopsy: not reported
Flow and timing	Flow: all participants in this subset of the PROPHESYS cohorts were included in the analysis Time between index test and biopsy: not reported
Comparative	Comparators: APRI
Notes	

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Low risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index test (FIB-4)			

Ferenci 2014 *(Continued)*

Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index test (Forns)			
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Unclear risk	

Fontaine 2009

Study characteristics	
Patient Sampling	Study design: cross-sectional diagnostic test accuracy study Sampling method: cohort-based Direction of data collection: not reported Country: France Inclusion criteria: HCV infected patients who were on haemodialysis or had received a renal transplant Exclusion criteria: not reported
Patient characteristics and setting	Centre details: single centre, Cochin Hospital, Paris, France Sample size: 110 Mean age: 58 Gender (% male): 60 Mean BMI: not reported Mean ALT: not reported

Fontaine 2009 *(Continued)*

	Special characteristics: 51 haemodialysis patients and 59 kidney recipients
Index tests	Test name(s): FIB-4 Threshold(s) used: 1.45
Target condition and reference standard(s)	Target condition(s): F2+, F3+, F4 Reference standard: liver biopsy Quality of liver biopsy: not reported
Flow and timing	Flow: all participants received the index test and none were excluded from the analysis Time between index test and biopsy: not reported
Comparative	Comparators: FibroTest
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		Unclear risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index test (FIB-4)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index test (Forns)			
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	Unclear		

Fontaine 2009 *(Continued)*

Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear	
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk
Are there concerns that the target condition as defined by the reference standard does not match the question?		Low concern
DOMAIN 4: Flow and timing		
Was there an appropriate interval between index test and reference standard?	Unclear	
Did all patients receive the same reference standard?	Yes	
Were all patients included in the analysis?	Yes	
Could the patient flow have introduced bias?		Unclear risk

Forns 2002a

Study characteristics	
Patient Sampling	Study design: cross-sectional diagnostic test accuracy study Method of sampling: cohort-based Direction of data collection: not reported Country: Spain Inclusion criteria: chronic HCV Exclusion criteria: older than 65 years, regular alcohol intake higher than 30 g/d, morbid obesity, HIV or HBV co-infection, active intravenous drug abuse, previous interferon treatment, liver transplantation, and clinical or ultrasonographic evidence of cirrhosis
Patient characteristics and setting	Centre details: single centre, Institut de Malalties Digestives, Hospital Clinic, Barcelona Sample size: 502 Mean age: 38 (validation set), 39 (estimation set) Gender (% male): 64 Mean BMI: not reported Mean ALT: 90 (validation set), 97 (estimation set)
Index tests	Test name(s): Forns index Threshold(s) used: 4.2, 6.9
Target condition and reference standard(s)	Target condition(s): F2+ Reference standard: liver biopsy Quality of liver biopsy: 6 portal tracts
Flow and timing	Flow: all participants received the index test and none were excluded from the analysis Time between index test and biopsy: same day
Comparative	Comparators: nil

Forns 2002a *(Continued)*

Notes	Data were available for both the estimation and the validation group. We considered these separately for the purposes of meta-analysis. The estimation and validation groups are referenced under Forns2002a and Forns 2002b, respectively.

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Low risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index test (FIB-4)			
DOMAIN 2: Index test (Forns)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and timing			
Was there an appropriate interval between index test and reference standard?	Yes		

Forns 2002a *(Continued)*

Did all patients receive the same reference standard?	Yes	
Were all patients included in the analysis?	Yes	
Could the patient flow have introduced bias?		Low risk

Forns 2002b

Study characteristics			
Patient Sampling	Study design: cross-sectional diagnostic test accuracy study Method of sampling: cohort-based Direction of data collection: not reported Country: Spain Inclusion criteria: chronic HCV Exclusion criteria: age greater than 65 years, regular alcohol intake higher than 30 g/d, morbid obesity, HIB or HBV co-infection, active intravenous drug abuse, previous interferon treatment, liver transplantation, and clinical or ultrasonographic evidence of cirrhosis		
Patient characteristics and setting	Centre details: single centre, Institut de Malalties Digestives, Hospital Clinic, Barcelona Sample size: 502 Mean age: 38 (validation set), 39 (estimation set) Gender (% male): 64 Mean BMI: not reported Mean ALT: 90 (validation set), 97 (estimation set)		
Index tests	Test name(s): Forns index Threshold(s) used: 4.2, 6.9		
Target condition and reference standard(s)	Target condition(s): F2+ Reference standard: liver biopsy Quality of liver biopsy: 6 portal tracts		
Flow and timing	Flow: all participants received the index test and none were excluded from the analysis Time between index test and biopsy: same day		
Comparative	Comparators: nil		
Notes	Data were available for both the estimation and the validation group. We considered these separately for the purposes of meta-analysis. The estimation and validation groups are referenced under Forns2002a and Forns 2002b, respectively.		
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	Yes		

Forns 2002b *(Continued)*

Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Low risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index test (FIB-4)			
DOMAIN 2: Index test (Forns)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Low risk	

Fouad 2018

Study characteristics	
Patient Sampling	Study design: cross-sectional diagnostic test accuracy study Method of sampling: cohort-based

Fouad 2018 *(Continued)*

	Direction of data collection: not reported Country: Egypt Inclusion criteria: chronic HCV Exclusion criteria: HBV, autoimmune hepatitis, decompensated cirrhosis, HCC, extrahepatic malignancy
Patient characteristics and setting	Centre details: single centre, Kasr Al-Aini Viral Hepatitis Center, Cairo University Sample size: 72 Mean age: 44 Gender (% male): 63 Mean BMI: 23.6 Mean ALT: 50
Index tests	Test name(s): FIB-4 Threshold(s) used: 1.5
Target condition and reference standard(s)	Target condition(s): F3+ Reference standard: liver biopsy Quality of liver biopsy: > 15 mm in length and including more than 11 portal tracts
Flow and timing	Flow: all participants received the index test and none were excluded from the analysis Time between index test and biopsy: not reported
Comparative	Comparators: ElastPQ, TE, APRI
Notes	

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Unclear risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index test (FIB-4)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	

Fouad 2018 *(Continued)*

Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index test (Forns)			
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Unclear risk	

Fujita 2018

Study characteristics	
Patient Sampling	Study design: cross-sectional diagnostic test accuracy study Method of sampling: cohort-based Direction of data collection: not reported Country: Japan Inclusion criteria: HCV infected patients who underwent liver biopsy examinations Exclusion criteria: HCC
Patient characteristics and setting	Centre details: single centre, Kagawa University, Japan Sample size: 122 Mean age: 53 Gender (% male): 67 Mean BMI: not reported Mean ALT: 98
Index tests	Test name(s): FIB-4 Threshold(s) used: 1.76, 3.73
Target condition and reference standard(s)	Target condition(s): F3+ Reference standard: liver biopsy Quality of liver biopsy: not reported

Fujita 2018 *(Continued)*

Flow and timing	Flow: all participants received the index test and none were excluded from the analysis Time between index test and biopsy: not reported
Comparative	Comparators: Wisteria floribunda agglutinin-positive Mac-2 binding protein (WFA-M2BP), Enhanced liver fibrosis (ELF) score, APRI
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Unclear risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index test (FIB-4)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index test (Forns)			
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and timing			

Fujita 2018 *(Continued)*

Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Unclear risk	

Gamil 2017

Study characteristics			
Patient Sampling	Study design: cross-sectional diagnostic test accuracy study Method of sampling: cohort-based Direction of data collection: retrospective Country: Egypt Inclusion criteria: chronic HCV treatment-naïve patients Exclusion criteria: other types of chronic liver disease, HCC		
Patient characteristics and setting	Centre details: single centre, Kasr Al-Ainy viral hepatitis treatment center in Cairo University Sample size: 652 Mean age: 41 Gender (% male): 69 Mean BMI: 28 Mean ALT: 56		
Index tests	Test name(s): FIB-4 Threshold(s) used: F2+ 1.05, F3+ 1.45, F4 2.00		
Target condition and reference standard(s)	Target condition(s): F2+, F3+, F4 Reference standard: liver biopsy Quality of liver biopsy: not reported		
Flow and timing	Flow: all participants received the index test and none were excluded from the analysis Time between index test and biopsy: LSM was done within 1 week of biopsy, but time between other index tests and biopsy not reported		
Comparative	Comparators: LSM, APRI, various novel scores that add alpha-fetoprotein (AFP) levels to known equations		
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	Yes		

Gamil 2017 *(Continued)*

Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Low risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index test (FIB-4)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index test (Forns)			
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Unclear risk	

Gorka-Dynysiewicz 2019

Study characteristics	
Patient Sampling	Study design: case-control diagnostic test accuracy study Method of sampling: case-control

Gorka-Dynysiewicz 2019 *(Continued)*

	Direction of data collection: not reported Country: Poland Inclusion criteria: cases – treatment-naïve chronic HCV as based on persistently increased ALT, anti-HCV and HCV-RNA positivity and liver histology features. Controls – admitted to the centre at the same time without above features Exclusion criteria: HBV, HIV, fatty liver, alcohol excess, intravenous drug use, autoimmune or congenital liver conditions, malignancies, immunosuppressants
Patient characteristics and setting	Centre details: single centre, Department of Infectious Diseases and Hepatology, Wroclaw Medical University Sample size: 138 Mean age: 55 Gender (% male): 60 Mean BMI: 22 Mean ALT: 64
Index tests	Test name(s): FIB-4 Threshold(s) used: 1.86
Target condition and reference standard(s)	Target condition(s): F2+ Reference standard: liver biopsy Quality of liver biopsy: not reported
Flow and timing	Flow: all participants received the index test and none were excluded from the analysis Time between index test and biopsy: same day
Comparative	Comparators: Pentraxin 3 (PTX3), transforming growth factor-1 (TGF-1), hyaluronic acid (HA), aspartate transaminase to platelet ratio index (APRI), gamma-glutamyl transpeptidase to platelet ratio (GPR), FibroScan
Notes	

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	No		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index test (FIB-4)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		

Gorka-Dynysiewicz 2019 *(Continued)*

If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index test (Forns)			
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Low risk	

Gorka-Dynysiewicz 2020

Study characteristics	
Patient Sampling	Study design: cross-sectional diagnostic test accuracy study Method of sampling: cohort-based Direction of data collection: not reported Country: Poland Inclusion criteria: treatment-naïve chronic HCV Exclusion criteria: not reported
Patient characteristics and setting	Centre details: single centre, Department of Infectious Diseases and Hepatology of Wroclaw Medical University Sample size: 242 (150 in training set, 92 in validation set) Mean age: training set 56, validation set 55 Gender (% male): training set 52, validation set 58 Mean BMI: training set 22.5, validation set 22.2 Mean ALT: training set 63, validation set 64
Index tests	Test name(s): FIB-4, Forns index

Gorka-Dynysiewicz 2020 *(Continued)*

	Threshold(s) used: FIB-4 1.86, Forns index 5.67
Target condition and reference standard(s)	Target condition(s): F2+ Reference standard: liver biopsy Quality of liver biopsy: not reported
Flow and timing	Flow: all participants received the index test and none were excluded from the analysis Time between index test and biopsy: same day
Comparative	Comparators: APRI, novel Pentra score model
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Unclear risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index test (FIB-4)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index test (Forns)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern

Gorka-Dynysiewicz 2020 *(Continued)*

DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Low risk	

Gounder 2018

Study characteristics	
Patient Sampling	Study design: cross-sectional diagnostic test accuracy study Method of sampling: cohort-based Direction of data collection: retrospective Country: USA Inclusion criteria: American Indian/Alaska Native adults aged 18 years and above with HCV infection who received health care through the Alaska Native Medical Center's Liver Diseases and Hepatitis Program Exclusion criteria: people with resolved HCV infection, uncertain HCV status, HIV/HBV co-infection
Patient characteristics and setting	Centre details: multicentre, Alaska Native Medical Center's Liver Diseases and Hepatitis Program Sample size: 457 Mean age: 46 Gender (% male): 54 Mean BMI: 137 Mean ALT: not reported Special characteristics: all participants were American Indian/Alaska Native adults
Index tests	Test name(s): FIB-4 Threshold(s) used: 1.45, 3.25
Target condition and reference standard(s)	Target condition(s): F2+, F3+ Reference standard: liver biopsy

Gounder 2018 *(Continued)*

	Quality of liver biopsy: not reported
Flow and timing	Flow: all participants received the index test and none were excluded from the analysis Time between index test and biopsy: within 1 month
Comparative	Comparators: APRI
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Low risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index test (FIB-4)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index test (Forns)			
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern

Gounder 2018 *(Continued)*

DOMAIN 4: Flow and timing		
Was there an appropriate interval between index test and reference standard?	Yes	
Did all patients receive the same reference standard?	Yes	
Were all patients included in the analysis?	Yes	
Could the patient flow have introduced bias?		Low risk

Guilabert 2010

Study characteristics	
Patient Sampling	Study design: cross-sectional diagnostic test accuracy Method of sampling: cohort-based Direction of data collection: retrospective Country: Spain Inclusion criteria: > 18 years old, consent, HCV serology, ALT increase in recent 6 months, satisfactory FibroScan Exclusion criteria: not reported
Patient characteristics and setting	Centre details: single centre, Hospital Clinico Universitario de Valladolid Sample size: 154 Mean age: 43.6 Gender (% male): 77 Mean BMI: not reported Mean ALT: 38.5 Special characteristics: 60% were infected with HIV
Index tests	Test name(s): FIB-4, Forns index Threshold(s) used: FIB-4 F1+ 1.4, F2+ 1.4, F3+ 2.1, F4 2.5; Forns index F1+ 5.5, F2+ 5.6, F3+ 5.7, F4 7.4
Target condition and reference standard(s)	Target condition(s): F1+, F2+, F3+, F4 Reference standard: liver biopsy Quality of liver biopsy: not reported
Flow and timing	Flow: all participants received the index test and none were excluded from the analysis Time between index test and biopsy: not reported
Comparative	Comparators: FibroScan, APRI
Notes	

Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient selection			

Guilabert 2010 *(Continued)*

Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Low risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index test (FIB-4)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index test (Forns)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Yes		

Guilabert 2010 *(Continued)*

Were all patients included in the analysis?	Yes
Could the patient flow have introduced bias?	Unclear risk

Gökan 2016

Study characteristics	
Patient Sampling	Study design: cross-sectional diagnostic test accuracy study Method of sampling: cohort-based Direction of data collection: retrospective Country: Turkey Inclusion criteria: treatment-naïve, positive serum HCV on PCR Exclusion criteria: decompensated liver disease, inadequate data, alcohol excess, co-infection with HCV or hepatitis D virus (HDV), autoimmune or metabolic disorders
Patient characteristics and setting	Centre details: single centre, Department of Hepatology-Gastroenterology of Türkiye Yüksek İhtisas Hospital Sample size: 120 Mean age: 51.7 Gender (% male): 57.5 Mean BMI: not reported Mean ALT: 48.5 Special characteristics: all had HCV genotype 1a
Index tests	Test name(s): FIB-4 Threshold(s) used: 1.45, 1.6, 3.25
Target condition and reference standard(s)	Target condition(s): F3+ Reference standard: liver biopsy Quality of liver biopsy: not reported
Flow and timing	Flow: all participants received the index test and none were excluded from the analysis Time between index test and biopsy: simultaneous to liver biopsy
Comparative	Comparators: age-platelet index (AP index), cirrhosis discriminant score (CDS), aspartate aminotransferase (AST)-alanine aminotransferase (ALT) ratio (AAR), AST-platelet ratio index (APRI), Göteborg University Cirrhosis Index (GUCI), FibroQ, King's score, platelet count
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		

Gökan 2016 *(Continued)*

Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Low risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index test (FIB-4)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index test (Forns)			
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Low risk	

Güzelbulut 2011

Study characteristics	
Patient Sampling	Study design: cross-sectional diagnostic test accuracy study Method of sampling: cohort-based Direction of data collection: retrospective Country: Turkey

Güzelbulut 2011 *(Continued)*

	Inclusion criteria: chronic HCV, treatment-naïve at time of biopsy, sufficient lab information within 3 months of biopsy Exclusion criteria: HIV/HBV co-infection, alcohol misuse, HCC, other liver diseases
Patient characteristics and setting	Centre details: single centre, Haydarpafla Numune Education and Research Hospital, Istanbul Sample size: 150 Mean age: 52 Gender (% male): 52 Mean BMI: not reported Mean ALT: 80
Index tests	Test name(s): FIB-4 and Forns index Threshold(s) used: FIB-4 F2+ 0.6, 1, F4 1.45, 3.25; Forns index F2+ 4.2, 6.9, F4 4.2, 6.9
Target condition and reference standard(s)	Target condition(s): F2+, F4 Reference standard: liver biopsy Quality of liver biopsy: not reported
Flow and timing	Flow: all participants received the index test and none were excluded from the analysis Time between index test and biopsy: within 3 months
Comparative	Comparators: APRI
Notes	

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Low risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index test (FIB-4)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	

Güzelbulut 2011 *(Continued)*

Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index test (Forns)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Low risk	

Hashem 2021

Study characteristics	
Patient Sampling	Study design: cross-sectional diagnostic test accuracy study Method of sampling: cohort-based Direction of data collection: not reported Country: Egypt Inclusion criteria: people with chronic HCV infections registered by the Egyptian National Committee for Control of Viral Hepatitis from January 2010 to December 2014 Exclusion criteria: previous interferon-based antiviral therapy
Patient characteristics and setting	Centre details: multicentre

Hashem 2021 *(Continued)*

	Sample size: 71,806 Mean age: 40 Gender (% male): 72 Mean BMI: 27 Mean ALT: 101
Index tests	Test name(s): F3+ Threshold(s) used: 1.45, 3.25
Target condition and reference standard(s)	Target condition(s): F3+ Reference standard: liver biopsy Quality of liver biopsy: not reported
Flow and timing	Flow: all participants received the index test and none were excluded from the analysis Time between index test and biopsy: not reported
Comparative	Comparators: APRI
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Unclear risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index test (FIB-4)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index test (Forns)			
DOMAIN 3: Reference standard			

Hashem 2021 *(Continued)*

Is the reference standard likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Unclear risk	

Holmberg 2013

Study characteristics	
Patient Sampling	Study design: cross-sectional diagnostic test accuracy study Method of sampling: cohort-based (chronic hepatitis C subpopulation of the Chronic Hepatitis Cohort Study) Direction of data collection: prospective Country: USA Inclusion criteria: confirmed chronic HCV Exclusion criteria: liver transplant patients
Patient characteristics and setting	Centre details: multicentre, Geisinger Health System, Danville, Pennsylvania; Henry Ford Health System, Detroit, Michigan; Kaiser Permanente Northwest, Portland, Oregon; Kaiser Permanente, Honolulu, Hawaii (CHeCS) Sample size: 10,473 Mean age: 50.1 Gender (% male): 60.8 Mean BMI: not reported Mean ALT: not reported
Index tests	Test name(s): FIB-4 Threshold(s) used: 1.81
Target condition and reference standard(s)	Target condition(s): F3+ Reference standard: liver biopsy Quality of liver biopsy: not reported
Flow and timing	Flow: all participants received the index test and none were excluded from the analysis Time between index test and biopsy: within 6 months

Holmberg 2013 *(Continued)*

Comparative	Comparators: AST/ALT ratio, APRI
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Low risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index test (FIB-4)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index test (Forns)			
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and timing			
Was there an appropriate interval between index test and reference standard?	No		
Did all patients receive the same reference standard?	Yes		

Holmberg 2013 *(Continued)*

Were all patients included in the analysis?	Yes
Could the patient flow have introduced bias?	High risk

Hseih 2012

Study characteristics	
Patient Sampling	Study design: cross-sectional diagnostic test accuracy study Method of sampling: cohort-based Direction of data collection: retrospective Country: Taiwan Inclusion criteria: chronic HCV infection confirmed by the presence of anti-HCV antibody by enzyme immunoassay methods Exclusion criteria: HIV/HBV co-infection, alcohol misuse, HCC, liver transplantation, metabolic liver disease, insufficient biopsy, recent anticoagulant use
Patient characteristics and setting	Centre details: single centre, Department of Gastroenterology, Chang Gung Memorial Hospital, Chiayi Sample size: 250 Mean age: 54.3 Gender (% male): 57 Mean BMI: not reported Mean ALT: 156
Index tests	Test name(s): FIB-4 Threshold(s) used: 1.45, 3.25
Target condition and reference standard(s)	Target condition(s): F3+ Reference standard: liver biopsy Quality of liver biopsy: not reported
Flow and timing	Flow: all participants received the index test and none were excluded from the analysis Time between index test and biopsy: less than 1 month
Comparative	Comparators: FibroQ, AAR, API, and Lok's model
Notes	

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Low risk	

Hseih 2012 *(Continued)*

Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index test (FIB-4)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index test (Forns)			
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Low risk	

Hsu 2019

Study characteristics	
Patient Sampling	Study design: cross-sectional diagnostic test accuracy study Method of sampling: cohort based Direction of data collection: prospective Country: Taiwan Inclusion criteria: > 18 years. Received acoustic radiation force impulse (ARFI) measurements within 4 weeks of liver histology study. Included autoimmune liver diseases and chronic hepatitis B patients (but reported HCV data separately)

Hsu 2019 *(Continued)*

	Exclusion criteria: alcoholism, co-infection with HBV, patients with ALT level > 5 times the upper limit of normal
Patient characteristics and setting	Centre details: single centre, Chang Gung Memorial Hospital Sample size: 606 Mean age: 52 Gender (% male): 59 Mean BMI: 26 Mean ALT: 64
Index tests	Test name(s): FIB-4 Threshold(s) used: F1+ 1.23, F2+ 1.69, F3+ 1.91, F4 2.14
Target condition and reference standard(s)	Target condition(s): F1+, F2+, F3+ F4 Reference standard: liver biopsy Quality of liver biopsy: not reported
Flow and timing	Flow: 96 participants who received both index test and biopsy were excluded from the analysis due to indeterminate liver stiffness measurements or excessively high ALT levels. Time between index test and biopsy: time between liver stiffness measurement and biopsy < 4 weeks but no comment on timing of other index tests
Comparative	Comparators: acoustic radiation force impulse (ARFI) imaging
Notes	

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Low risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index test (FIB-4)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern

Hsu 2019 *(Continued)*

DOMAIN 2: Index test (Forns)			
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	No		
Could the patient flow have introduced bias?		High risk	

Iacobellis 2005

Study characteristics	
Patient Sampling	Study design: cross-sectional diagnostic test accuracy study Method of sampling: cohort-based Direction of data collection: not reported Country: Italy Inclusion criteria: treatment-naïve chronic HCV infection, elevated aminotransferase levels for > 6 months, detectable levels of HCV RNA, and compatible hepatic histology Exclusion criteria: other causes of liver disease
Patient characteristics and setting	Centre details: single centre Hospital S. De Bellis, Castellana Grotte, Matera, Italy Sample size: 1252 Mean age: 54.4 Gender (% male): 57 Mean BMI: not reported Mean ALT: 112
Index tests	Test name(s): Forns Threshold(s) used: 6.9
Target condition and reference standard(s)	Target condition(s): F3+ Reference standard: liver biopsy Quality of liver biopsy: minimum of 5 portal tracts

Iacobellis 2005 *(Continued)*

Flow and timing	Flow: all participants received the index test and none were excluded from the analysis Time between index test and biopsy: same day
Comparative	Comparators: AST/ALT ratio, platelet count, APRI
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Low risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index test (FIB-4)			
DOMAIN 2: Index test (Forns)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and timing			

Iacobellis 2005 *(Continued)*

Was there an appropriate interval between index test and reference standard?	Yes	
Did all patients receive the same reference standard?	Yes	
Were all patients included in the analysis?	Yes	
Could the patient flow have introduced bias?		Low risk

Ikatura 2021

Study characteristics			
Patient Sampling	Study design: cohort, diagnostic test accuracy study Method of sampling: cohort-based Direction of data collection: prospective and retrospective Country: Japan Inclusion criteria: chronic hepatitis B, chronic hepatitis C, liver histology Exclusion criteria: other liver disease (i.e. complication of liver cancer), patients with insufficient data		
Patient characteristics and setting	Centre details: multicentre, 11 referral hospitals in Japan Sample size: 1029 people with HCV Mean age: 62.6 Gender (% male): 44.1% Mean BMI: not reported Mean ALT: 65.5		
Index tests	Test name(s): FIB-4 Threshold(s) used: F2+ 2.70, 3.26, 3.47, F4 3.65, 3.61, 4.32		
Target condition and reference standard(s)	Target condition(s): F3+ and F4 Reference standard: liver biopsy Quality of liver biopsy: not reported		
Flow and timing	Flow: all participants received the index test and none were excluded from the analysis Time between index test and biopsy: unclear		
Comparative	Comparators: FIB-4, APRI		
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	Yes		

Ikatura 2021 *(Continued)*

Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Unclear risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index test (FIB-4)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index test (Forns)			
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Unclear risk	

Kamphues 2010

Study characteristics	
Patient Sampling	Study design: cross-sectional diagnostic test accuracy study Method of sampling: cohort-based Direction of data collection: prospective Country: Germany

Kamphues 2010 *(Continued)*

	Inclusion criteria: liver transplation after final liver disease caused by hepatitis C or alcoholic disease Exclusion criteria: obesity, inadequate liver biopsy, contraindications for liver biopsy, acute rejection, chronic rejection, ascites
Patient characteristics and setting	Centre details: single centre, Universitatsklinikum Charite, Humboldt-Universitat, Berlin, Germany Sample size: 101 in HCV cohort (further 57 in alcoholic cirrhosis cohort, with data reported separately) Mean age: 51.7 Gender (% male): 64.9% Mean BMI: 25 Mean ALT: 51.8 Special characteristics: liver transplant recipients
Index tests	Test name(s): FIB-4 Threshold(s) used: 2.8 for F ≥ 2; 4.44 for F = 4.
Target condition and reference standard(s)	Target condition(s): F2+ and F4 Reference standard: liver biopsy Quality of liver biopsy: minimum length of 1.5 cm, no mention of portal tract number
Flow and timing	Flow: 2 participants were excluded for inadequate liver biopsy, 5 were excluded for obesity Time between index test and biopsy: the same time or within 2 days
Comparative	Comparators: APRI, FibroScan
Notes	

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index test (FIB-4)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		

Kamphues 2010 *(Continued)*

Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index test (Forns)			
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	No		
Could the patient flow have introduced bias?		High risk	

Kitajima 2016

Study characteristics	
Patient Sampling	Study design: cross-sectional diagnostic test accuracy study Method of sampling: cohort-based Direction of data collection: retrospective Country: Japan Inclusion criteria: HCV-related end-stage liver disease Exclusion criteria: HBV co-infection, pretransplant splenectomy, splenectomy not performed
Patient characteristics and setting	Centre details: single centre Sample size: 110 Mean age: 59 Gender (% male): 53.6% Mean BMI: 23.0 Mean ALT: not reported Special characteristics: participants were living-donor liver transplantation patients who had undergone concomitant splenectomy
Index tests	Test name(s): FIB-4

Kitajima 2016 *(Continued)*

	Threshold(s) used: 2.20 (optimal cut-off for significant fibrosis, Youden)
Target condition and reference standard(s)	Target condition(s): F2+ Reference standard: liver biopsy Quality of liver biopsy: minimal acceptable size of liver biopsy was 15 mm, portal tracts not reported
Flow and timing	Flow: all enroled participants received the index test and none were excluded from the analysis Time between index test and biopsy: 1 week
Comparative	Comparators: AST/ALT ratio, APRI, Age-platelet index
Notes	

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		Unclear risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index test (FIB-4)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index test (Forns)			
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		

Kitajima 2016 *(Continued)*

Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk
Are there concerns that the target condition as defined by the reference standard does not match the question?		Low concern
DOMAIN 4: Flow and timing		
Was there an appropriate interval between index test and reference standard?	Yes	
Did all patients receive the same reference standard?	Yes	
Were all patients included in the analysis?	Yes	
Could the patient flow have introduced bias?		Low risk

Koller 2014

Study characteristics	
Patient Sampling	Study design: cross-sectional diagnostic test accuracy study Method of sampling: cohort-based Direction of data collection: prospective Country: Slovakia Inclusion criteria: people with HCV infection undergoing liver biopsy before first antiviral treatment Exclusion criteria: missing laboratory data, insufficient biopsy samples
Patient characteristics and setting	Centre details: multicentre: 4 centres Sample size: 104 (52 in training group, 52 in validation group) Mean age: 35.37 in training group, 32.67 in validation group Gender (% male): 73.1% in training group, 67.3% in validation group Mean BMI: 25.8 in training group, 25.9 in validation group Mean ALT: 1.4 times the ULN in training group, 2.2 times the ULN in validation group
Index tests	Test name(s): both FIB-4 and Forns index Threshold(s) used: for Forns index: 4.2 and 6.9. For FIB-4: 0.6 and 1.0
Target condition and reference standard(s)	Target condition(s): Ishak fibrosis stage greater than 2 (significant liver fibrosis) Reference standard: liver biopsy Quality of liver biopsy: specimen > 10 mm or the sum of fragment lengths > 10 mm
Flow and timing	Flow: 9 cases (just under 10% of total included participants) were excluded from the analysis: 7 because of missing laboratory data and 2 cases because of insufficient biopsy samples Time between index test and biopsy: not reported
Comparative	Comparators: APRI, noninvasive fibrosis score (NFS)

Koller 2014 *(Continued)*

Notes

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Low risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index test (FIB-4)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index test (Forns)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	

Koller 2014 *(Continued)*

Are there concerns that the target condition as defined by the reference standard does not match the question?		Low concern
DOMAIN 4: Flow and timing		
Was there an appropriate interval between index test and reference standard?	Yes	
Did all patients receive the same reference standard?	Yes	
Were all patients included in the analysis?	No	
Could the patient flow have introduced bias?		High risk

Ladero 2010

Study characteristics	
Patient Sampling	Study design: cross-sectional diagnostic test accuracy study Method of sampling: cohort-based Direction of data collection: retrospective Country: Spain Inclusion criteria: chronic hepatitis C non-treated patients, at least a known positive result for anti-HCV > 6 months before the date of the biopsy and a positive determination of serum HCV RNA < 2 months before the liver biopsy Exclusion criteria: current ethanol abuse (> 40 g/day), active hepatitis B virus infection, HIV antibody (Ab)-positivity, coexistence of genetic or autoimmune liver disease
Patient characteristics and setting	Centre details: single centre, Hospital Clínico San Carlos, Universidad Complutense, Madrid Sample size: 429 (two distinct cohorts: 243 with < F2, 186 with F2+ Mean age: 43.0 (in null-low fibrosis group, F < 2) and 47.1 (in significant-advanced fibrosis group, F2+) Gender (% male): 57.2 (in null-low fibrosis group, F < 2) and 61.3 (in significant-advanced fibrosis group, F2+) Mean BMI: not reported Mean ALT: 87 (in null-low fibrosis group, F < 2) and 127 (in significant-advanced fibrosis group, F2+)
Index tests	Test name(s): both FIB-4 and Forns index Threshold(s) used: for Forns index: 4.2. For FIB-4: 1.45
Target condition and reference standard(s)	Target condition(s): F2+ Reference standard: liver biopsy Quality of liver biopsy: specimen length of 10 mm or more (with subgroup analysis on participants with specimens 15 mm or more). However, no mention of minimum number of portal tracts required.
Flow and timing	Flow: all participants received the index test and were included in the analysis Time between index test and biopsy: same day
Comparative	Comparators: King's, GUCI, APRI x ln ferritin/ln cholesterol, King's x ln ferritin/ln cholesterol

Ladero 2010 *(Continued)*

Notes

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Low risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index test (FIB-4)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index test (Forns)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	

Ladero 2010 *(Continued)*

Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Low risk	

Leroy 2007

Study characteristics	
Patient Sampling	Study design: cross-sectional diagnostic test accuracy study Method of sampling: cohort-based Direction of data collection: prospective Country: France Inclusion criteria: anti-HCV positive by ELISA, detectable serum HCV-RNA by PCR, elevated ALT serum levels Exclusion criteria: co-infection with HIV or HBV, other causes of liver disease, alcohol consumption higher than 30 g/day, hepatocellular carcinoma, Gilbert disease, chronic haemolysis, inflammatory syndrome and previous antiviral treatment
Patient characteristics and setting	Centre details: single centre, CHU de Grenoble, France Sample size: 180 Mean age: 43.7 Gender (% male): 62.2 Mean BMI: not reported Mean ALT: 73
Index tests	Test name(s): Forns index Threshold(s) used: 4.20 and 6.90
Target condition and reference standard(s)	Target condition(s): F2+, F3+. Reference standard: liver biopsy Quality of liver biopsy: biopsy length was greater than 15 mm in 161 (89.4%) and greater than 25 mm in 81 (45.0%) participants. Minimum length and number of portal tracts not reported.
Flow and timing	Flow: all participants received the index test and none were excluded from the analysis Time between index test and biopsy: the same day
Comparative	Comparators: Fibrometer, APRI, MP3, FibroTest, Hepascore
Notes	
Methodological quality	

Leroy 2007 *(Continued)*

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Low risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index test (FIB-4)			
DOMAIN 2: Index test (Forns)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Low risk	

Loko 2008

Study characteristics			
Patient Sampling	Study design: cross-sectional diagnostic test accuracy study Method of sampling: cohort-based Direction of data collection: retrospective Country: France Inclusion criteria: > 18 years, positive serum antibodies to HCV by means of a second- or third-generation HCV enzyme-linked immunosorbent assay and detectable serum HCV RNA Exclusion criteria: HBV co-infection, other known causes of liver disease, and alcohol intake of more than 50 g/day		
Patient characteristics and setting	Centre details: single centre, University of Bordeaux hospital Sample size: 200 Mean age: 39.8 Gender (% male): 67.0 Mean BMI: not reported Mean ALT: not reported Special characteristics: all participants were HCV-HIV co-infected		
Index tests	Test name(s): both FIB-4 and Forns index Threshold(s) used: FIB-4 F2+ 0.6, 1.0; FIB-4 F3+/F4 1.45, 3.25; Forns index F2+ 4.2, 6.0		
Target condition and reference standard(s)	Target condition(s): F2+, F3+ and F4. Reference standard: liver biopsy Quality of liver biopsy: not reported		
Flow and timing	Flow: all participants received the index test and none were excluded from the analysis Time between index test and biopsy: the same day or within 1 month		
Comparative	Comparators: APRI		
Notes			

Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Low risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern

Loko 2008 *(Continued)*

DOMAIN 2: Index test (FIB-4)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index test (Forns)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Low risk	

Macías 2006

Study characteristics			

Macías 2006 *(Continued)*

Patient Sampling	Study design: cross-sectional diagnostic test accuracy study Method of sampling: cohort-based Direction of data collection: retrospective Country: Spain Inclusion criteria: people co-infected with HIV and HCV who had undergone liver biopsy, regardless of levels of transaminases Exclusion criteria: positive hepatitis B surface antigen, other causes of liver disease (autoimmune, tumoral, biliary, or vascular associated liver disease) and prior anti-HCV therapy.
Patient characteristics and setting	Centre details: multicentre: 5 centres Sample size: 398 Mean age: 37 Gender (% male): 83% Mean BMI: not reported Mean ALT: 83 Special characteristics: all participants were HCV-HIV co-infected
Index tests	Test name(s): Forns index Threshold(s) used: 4.2 and 6.9
Target condition and reference standard(s)	Target condition(s): F2+, F4 Reference standard: liver biopsy Quality of liver biopsy: minimum liver biopsy length of 10 mm was required, no mention of number of portal tracts
Flow and timing	Flow: 263 participants (from the original 398 with both index and reference tests) were included in the analysis. Time between index test and biopsy: the same day and within 1 month
Comparative	Comparators: APRI, Saadeh, AST/ALT ratio, platelet count, Bonacini
Notes	

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Low risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index test (FIB-4)			
DOMAIN 2: Index test (Forns)			

Macías 2006 *(Continued)*

Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	No		
Could the patient flow have introduced bias?		High risk	

Maheshwari 2013

Study characteristics	
Patient Sampling	Study design: cross-sectional diagnostic test accuracy study Method of sampling: cohort-based Direction of data collection: retrospective Country: Georgia (USA) Inclusion criteria: chronic hepatitis C, with liver biopsies performed and laboratory data within 6 months of liver biopsy Exclusion criteria: not reported
Patient characteristics and setting	Centre details: single centre, Emory University, Atlanta, USA Sample size: 204 Mean age: not reported Gender (% male): not reported Mean BMI: not reported Mean ALT: not reported
Index tests	Test name(s): FIB-4

Maheshwari 2013 *(Continued)*

	Threshold(s) used: 1.30, 3.23
Target condition and reference standard(s)	Target condition(s): F3+ Reference standard: liver biopsy Quality of liver biopsy: not reported
Flow and timing	Flow: all participants received the index test and none were excluded from the analysis Time between index test and biopsy: within 6 months
Comparative	Comparators: none
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Low risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index test (FIB-4)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index test (Forns)			
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	

Maheshwari 2013 *(Continued)*

Are there concerns that the target condition as defined by the reference standard does not match the question?		Low concern
DOMAIN 4: Flow and timing		
Was there an appropriate interval between index test and reference standard?	No	
Did all patients receive the same reference standard?	Yes	
Were all patients included in the analysis?	Yes	
Could the patient flow have introduced bias?		High risk

Martinez 2011

Study characteristics	
Patient Sampling	Study design: prospective cohort study Method of sampling: cohort-based Direction of data collection: not reported Country: Spain Inclusion criteria: chronic hepatitis C established by HCV-RNA presence using polymerase chain reaction assays. All participants underwent a pretreatment liver biopsy within 6 months prior to the initiation of therapy. Exclusion criteria: HIV and/or HBV co-infection, other causes of chronic liver disease
Patient characteristics and setting	Centre details: single centre Sample size: 340 (cohort was assessed pre- and post-treatment, with data reported separately) Mean age: 47.7 Gender (% male): 64% Mean BMI: 25.4 Mean ALT: 2.94 times the ULN
Index tests	Test name(s): both Forns index and FIB-4 Threshold(s) used: for Forns index: 4.2 and 6.9 (to predict F2+). For FIB-4: 1.45 and 3.25 (to predict F3+)
Target condition and reference standard(s)	Target condition(s): F2+, F3+, F4 Reference standard: liver biopsy Quality of liver biopsy: minimum 6 portal tracts and 15 mm length
Flow and timing	Flow: all participants received the index test and none were excluded from the analysis Time between index test and biopsy: not reported
Comparative	Comparators: APRI, ELF
Notes	
Methodological quality	

Martinez 2011 *(Continued)*

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Low risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index test (FIB-4)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index test (Forns)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and timing			

Martinez 2011 *(Continued)*

Was there an appropriate interval between index test and reference standard?	No		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Unclear risk	

Matsuura 2018

Study characteristics			
Patient Sampling	Study design: cross-sectional diagnostic test accuracy study Method of sampling: cohort-based Direction of data collection: not reported Country: Japan Inclusion criteria: chronic hepatitis C and liver biopsy performed Exclusion criteria: HBV and/or HIV co-infection, other liver disease (such as autoimmune hepatitis, non-alcoholic steatohepatitis, and primary biliary cirrhosis)		
Patient characteristics and setting	Centre details: single centre Sample size: 84 Mean age: not reported Gender (% male): 57.14% Mean BMI: not reported Mean ALT: 37		
Index tests	Test name(s): FIB-4 Threshold(s) used: 2.745		
Target condition and reference standard(s)	Target condition(s): F4 Reference standard: liver biopsy Quality of liver biopsy: not reported		
Flow and timing	Flow: all participants received the index test and none were excluded from the analysis Time between index test and biopsy: same day		
Comparative	Comparators: circulating let-7a-5p levels, APRI, Mac-2 binding protein glycan isomer (M2BPGi), FibroScan		
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	Yes		

Matsuura 2018 *(Continued)*

Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Unclear risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index test (FIB-4)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index test (Forns)			
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Low risk	

Mossong 2011

Study characteristics	
Patient Sampling	Study design: cross-sectional diagnostic test accuracy study Method of sampling: cohort-based Direction of data collection: retrospective (data obtained from stored serum samplings and available liver biopsy)

Mossong 2011 *(Continued)*

	Country: Belgium Inclusion criteria: chronic hepatitis C and liver biopsy performed Exclusion criteria: not reported
Patient characteristics and setting	Centre details: single centre, Centre Hospitalier de Luxembourg Sample size: 186 Mean age: 39 Gender (% male): 64% Mean BMI: not reported Mean ALT: 186
Index tests	Test name(s): Forns index Threshold(s) used: cut-offs tested for Forns index in significant fibrosis (F2+): 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9. Cut-offs tested for Forns index in advanced fibrosis (F3+): 1.00, 2.00, 3.00, 4.00, 4.20, 5.00, 6.00, 6.90
Target condition and reference standard(s)	Target condition(s): F2+, F3+ Reference standard: liver biopsy Quality of liver biopsy: not reported
Flow and timing	Flow: all participants received the index test and none were excluded from the analysis Time between index test and biopsy: unclear; liver biopsy and blood test samples were stored "contemporarily"
Comparative	Comparators: FibroTest
Notes	

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Low risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index test (FIB-4)			
DOMAIN 2: Index test (Forns)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		

Mossong 2011 *(Continued)*

Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Unclear risk	

Nan 2019

Study characteristics	
Patient Sampling	Study design: cross-sectional diagnostic test accuracy study Method of sampling: cohort-based Direction of data collection: not reported Country: China Inclusion criteria: chronic hepatitis C, diagnosed on the basis of positive tests for serum antibodies against HCV and the presence of HCV RNA in the plasma in the previous 6 months. Eligible patients were > 18 years of age Exclusion criteria: presence of decompensated cirrhosis, co-infection with human immunodeficiency virus, hepatitis A, B or D virus infection, other causes of chronic liver disease or comorbidities precluding interferon therapy
Patient characteristics and setting	Centre details: single centre, Third Hospital of Hebei Medical University (Shijiazhuang, China) Sample size: 112 Mean age: 43 (for F < 2), 52.3 (for F2 and F3), 55.2 (for F4) Gender (% male): 50% Mean BMI: 24.2 (for F < 2), 25.4 (for F2 and F3), 25.4 (for F4) Mean ALT: 31.5 (for F < 2), 33.0 (for F2 and F3), 38.4 (for F4)

Nan 2019 *(Continued)*

Index tests	Test name(s): FIB-4 Threshold(s) used: 2.49
Target condition and reference standard(s)	Target condition(s): F2+, F4 Reference standard: liver biopsy Quality of liver biopsy: not reported
Flow and timing	Flow: all participants received the index test and none were excluded from the analysis Time between index test and biopsy: not reported
Comparative	Comparators: APRI, Liver Stiffness Measurement, Serum miR-1273g-3p
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Unclear risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index test (FIB-4)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index test (Forns)			
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		

Nan 2019 *(Continued)*

Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk
Are there concerns that the target condition as defined by the reference standard does not match the question?		Low concern
DOMAIN 4: Flow and timing		
Was there an appropriate interval between index test and reference standard?	Unclear	
Did all patients receive the same reference standard?	Yes	
Were all patients included in the analysis?	Yes	
Could the patient flow have introduced bias?		Unclear risk

Omran 2018

Study characteristics	
Patient Sampling	Study design: cross-sectional diagnostic test accuracy study Method of sampling: cohort-based Direction of data collection: not reported Country: Egypt Inclusion criteria: treatment-naive chronic hepatitis C patients Exclusion criteria: people with previous anti-HCV therapy, HBV co-infection, decompensated liver disease, hepatocellular carcinoma, body mass index > 30, and presence of absolute contraindication for liver biopsy
Patient characteristics and setting	Centre details: single centre, Faculty of Medicine, Cairo University, Cairo, Egypt Sample size: 100 Mean age: 46.8 Gender (% male): 57% Mean BMI: 25.6
Index tests	Test name(s): FIB-4 Threshold(s) used: 1.7
Target condition and reference standard(s)	Target condition(s): F2+ Reference standard: liver biopsy Quality of liver biopsy: not reported
Flow and timing	Flow: all participants received the index test and none were excluded from the analysis Time between index test and biopsy: not reported
Comparative	Comparators: FibroScan, AAR, APRI, AP index, FibroQ, CDS, King's score, GUCI, Combined AP index and FibroQ
Notes	
Methodological quality	

Omran 2018 *(Continued)*

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	No		
Could the selection of patients have introduced bias?		High risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index test (FIB-4)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index test (Forns)			
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Unclear risk	

Ozel 2015

Study characteristics	
Patient Sampling	Study design: cross-sectional diagnostic test accuracy study Method of sampling: cohort-based Direction of data collection: retrospective Country: Turkey Inclusion criteria: chronic hepatitis C patients who underwent a percutaneous liver biopsy. All participants were positive for anti-HCV antibodies for at least 6 months, with HCV-RNA levels higher than 104 IU/L Exclusion criteria: concomitant chronic liver disease, decompensated cirrhosis and hepatocellular carcinoma, previous interferon treatment, history of alcohol use (> 20 g/day), and other conditions that may affect liver function tests and platelets
Patient characteristics and setting	Centre details: single centre, Kayseri Training and Research Hospital Department of Gastroenterology Sample size: 137 Mean age: 53.5 Gender (% male): 39% Mean BMI: not reported Mean ALT: 60
Index tests	Test name(s): FIB-4 Threshold(s) used: 1.20 and 1.85 (to determine significant fibrosis); 1.50 and 2.65 (to determine cirrhosis)
Target condition and reference standard(s)	Target condition(s): F3+ and F4 Reference standard: liver biopsy Quality of liver biopsy: minimum 6 portal tracts
Flow and timing	Flow: all participants received the index test and none were excluded from the analysis Time between index test and biopsy: the same day or within 1 week
Comparative	Comparators: AST/ALT ratio, APRI, age-platelet index, GUCI, PAPAS (Platelet/Age/Phosphatase/AFP/AST index), CDS
Notes	

Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Low risk	

Ozel 2015 *(Continued)*

Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index test (FIB-4)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index test (Forns)			
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Low risk	

Paranaguá-Vezozzo 2017

Study characteristics	
Patient Sampling	Study design: cross-sectional diagnostic test accuracy study Method of sampling: cohort-based Direction of data collection: not reported Country: Brazil Inclusion criteria: recent HCV PCR, clinical chronic HCV, representative liver biopsy within 1 month of index test

Paranaguá-Vezozzo 2017 *(Continued)*

	Exclusion criteria: < 18 years old, HBV/HIV co-infection, cholestasis, non-alcoholic steatohepatitis, autoimmune hepatitis, haemochromatosis, Wilson's disease, decompensation
Patient characteristics and setting	Centre details: single centre, Hepatology Outpatient Center of Hospital das Clínicas, University of São Paulo School of Medicine, Brazil Sample size: 81 Mean age: 51 Gender (% male): 49.4% Mean BMI: 26.5 (median) Mean ALT: 50 (median)
Index tests	Test name(s): FIB-4 Threshold(s) used: 1.47 (for F2+); 2.0 (for F3+); 3.91 (for F4)
Target condition and reference standard(s)	Target condition(s): F2+, F3+, F4 Reference standard: liver biopsy Quality of liver biopsy: minimum of 10 portal spaces
Flow and timing	Flow: all participants received the index test and none were excluded from the analysis Time between index test and biopsy: within 30 days
Comparative	Comparators: APRI, FibroScan, ARFI
Notes	

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Low risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index test (FIB-4)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern

Paranaguá-Vezozzo 2017 *(Continued)*

DOMAIN 2: Index test (Forns)			
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Low risk	

Patel 2017

Study characteristics	
Patient Sampling	Study design: cross-sectional diagnostic test accuracy study Method of sampling: cohort-based Direction of data collection: retrospective Country: Houston (USA) Inclusion criteria: age > 18 years, diagnosis of chronic hepatitis C and end-stage renal disease (ESRD) Exclusion criteria: not reported
Patient characteristics and setting	Centre details: single centre Sample size: 50 Mean age: 56.66 Gender (% male): 68% Mean BMI: not reported Mean ALT: not reported Special characteristics: all participants had been diagnosed with end-stage kidney disease and were on haemodialysis
Index tests	Test name(s): FIB-4 Threshold(s) used: 3.25 (for F3+)
Target condition and reference standard(s)	Target condition(s): F3+ Reference standard: liver biopsy Quality of liver biopsy: not reported

Patel 2017 *(Continued)*

Flow and timing	Flow: all participants received the index test and none were excluded from the analysis Time between index test and biopsy: not reported
Comparative	Comparators: APRI
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Low risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index test (FIB-4)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index test (Forns)			
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and timing			

Patel 2017 *(Continued)*

Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Unclear risk	

Portilla 2009

Study characteristics			
Patient Sampling	Study design: cross-sectional diagnostic test accuracy study Method of sampling: cohort-based Direction of data collection: not reported Country: Spain Inclusion criteria: chronic hepatitis C and liver biopsy Exclusion criteria: pregnant or lactating women, people with coagulation disorders, severe psychiatric or neurological disease, anatomical abnormalities, or focal liver lesions		
Patient characteristics and setting	Centre details: multicentre: two prisons in Spain Sample size: 165 Mean age: 36.3 Gender (% male): 98.2% Mean BMI: not reported Mean ALT: 76 Special characteristics: recruited exclusively from penitentiaries		
Index tests	Test name(s): FIB-4 Threshold(s) used: 1.27 (for F2+); 1.30 (for F3+)		
Target condition and reference standard(s)	Target condition(s): F2+, F3+ Reference standard: liver biopsy Quality of liver biopsy: 6 or more portal tracts		
Flow and timing	Flow: all participants received the index test and none were excluded from the analysis Time between index test and biopsy: not reported		
Comparative	Comparators: APRI		
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		

Portilla 2009 *(Continued)*

Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Low risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index test (FIB-4)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index test (Forns)			
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Unclear risk	

Qian 2019

Study characteristics	
Patient Sampling	Study design: cross-sectional diagnostic test accuracy study Method of sampling: cohort-based Direction of data collection: retrospective Country: China

Qian 2019 *(Continued)*

	Inclusion criteria: chronic hepatitis C, liver biopsy to assess the progression of liver disease, treatment-naive patients Exclusion criteria: acute hepatitis C, co-infection with other hepatitis viruses, NAFLD, NAFLD, autoimmune liver disease, alcohol- or drug-related liver disease
Patient characteristics and setting	Centre details: Beijing Youan Hospital Sample size: 120 (cohort B, cohort A not used in our study) Mean age: 51.33 Gender (% male): 47.5% Mean BMI: 22.34 Mean ALT: 37.6 Special characteristics: HCV infection through regular plasma donations with repeated blood retransfusions between 1992 and 1995.
Index tests	Test name(s): FIB-4 Threshold(s) used: regarding cohort B: 3.25 (for F3+)
Target condition and reference standard(s)	Target condition(s): F3+ Reference standard: liver biopsy Quality of liver biopsy: not reported
Flow and timing	Flow: all participants received the index test and none were excluded from the analysis Time between index test and biopsy: not reported
Comparative	Comparators: serum golgi protein (GP)73, APRI
Notes	

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Unclear risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index test (FIB-4)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	

Qian 2019 *(Continued)*

Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index test (Forns)			
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Unclear risk	

Ramzy 2021

Study characteristics	
Patient Sampling	Study design: cross-sectional diagnostic test accuracy study Method of sampling: cohort-based Direction of data collection: not reported Country: Egypt Inclusion criteria: adults aged 18 to 60 years, evidence of chronic HCV infection, positive HCV antibody and HCV-RNA PCR confirmed by pathological evidence of chronic hepatitis, and naive to antiviral therapy Exclusion criteria: contraindications to liver biopsy, decompensated liver disease, hepatocellular carcinoma, or hepatitis B virus (HBV) co-infection
Patient characteristics and setting	Centre details: single centre Sample size: 197 (group 1: 100 participants with insignificant fibrosis; group 2: 97 participants with significant fibrosis) Mean age: 39.7 (group 1), 46.6 (group 2) Gender (% male): 72% (group 1), 66% (group 2) Mean BMI: 31.33 (group 1), 35.11 (group 2) Mean ALT: 44.5 (group 1), 74.49 (group 2)
Index tests	Test name(s): FIB-4

Ramzy 2021 *(Continued)*

	Threshold(s) used: 1.29 and 2.91
Target condition and reference standard(s)	Target condition(s): F2+, F3+ Reference standard: liver biopsy Quality of liver biopsy: not reported
Flow and timing	Flow: all participants received the index test and none were excluded from the analysis Time between index test and biopsy: not reported
Comparative	Comparators: APRI, APA, Liver Stiffness Measurement
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Unclear risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index test (FIB-4)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index test (Forns)			
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	

Ramzy 2021 *(Continued)*

Are there concerns that the target condition as defined by the reference standard does not match the question?		Low concern
DOMAIN 4: Flow and timing		
Was there an appropriate interval between index test and reference standard?	Unclear	
Did all patients receive the same reference standard?	Yes	
Were all patients included in the analysis?	Yes	
Could the patient flow have introduced bias?		Unclear risk

Schmid 2015

Study characteristics	
Patient Sampling	Study design: cross-sectional diagnostic test accuracy study Method of sampling: cohort-based Direction of data collection: prospective Country: Switzerland Inclusion criteria: HIV-positive participants with chronic HCV infection (detectable HCV-RNA by PCR for at least six months), liver biopsy Exclusion criteria: not reported
Patient characteristics and setting	Centre details: multicentre as described by the Swiss HIV Cohort Study Sample size: 105 Mean age: 43 Gender (% male): 77.1% Mean BMI: not reported Mean ALT: not reported Special characteristics: all participants were HIV-HCV co-infected
Index tests	Test name(s): FIB-4 Threshold(s) used: F2+ 1.45, 2.63; F4 1.45, 3.25, 1.94
Target condition and reference standard(s)	Target condition(s): F2+, F4. Reference standard: liver biopsy Quality of liver biopsy: three in four biopsies (but not all) were at least 20 mm in length and had more than 11 portal tracts. However, the quality of the liver biopsies in the remaining quarter remains unknown.
Flow and timing	Flow: all participants received the index test and none were excluded from the analysis Time between index test and biopsy: same day
Comparative	Comparators: APRI, FibroTest, hyaluronic acid, HepaScore, ELF
Notes	
Methodological quality	

Schmid 2015 *(Continued)*

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Low risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index test (FIB-4)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index test (Forns)			
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Low risk	

Schmoyer 2020

Study characteristics	
Patient Sampling	Study design: cross-sectional diagnostic test accuracy study Method of sampling: cohort-based Direction of data collection: retrospective Country: USA Inclusion criteria: haemodialysis-dependent and tested positive for HCV RNA Exclusion criteria: hepatitis B, non-alcoholic fatty liver disease, primary sclerosing cholangitis, primary biliary cholangitis, active alcohol abuse, active malignancy, heart failure, use of immuno-suppression, previous liver transplantation, people younger than 18 years
Patient characteristics and setting	Centre details: single centre Sample size: 139 Mean age: 52.8 Gender (% male): 76.3% Mean BMI: 27.4 Mean ALT: 36.2 Special characteristics: all participants were HCV-infected and dialysis-dependent end-stage renal failure
Index tests	Test name(s): FIB-4 Threshold(s) used: F3+ 1.23, 2.13, 3.25
Target condition and reference standard(s)	Target condition(s): F3+ Reference standard: liver biopsy Quality of liver biopsy: not reported
Flow and timing	Flow: all participants received the index test and none were excluded from the analysis Time between index test and biopsy: within 10 weeks of liver biopsy
Comparative	Comparators: AST/ALT ratio, APRI, Fibrosis index score, King's score
Notes	

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Unclear risk	

Schmoyer 2020 *(Continued)*

Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index test (FIB-4)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index test (Forns)			
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Low risk	

Sebastiani 2008a

Study characteristics	
Patient Sampling	Study design: cross-sectional diagnostic test accuracy study Method of sampling: cohort-based Direction of data collection: not reported Country: Italy Inclusion criteria: compensated chronic liver disease and diagnostic percutaneous liver biopsy.

Sebastiani 2008a *(Continued)*

	Exclusion criteria: co-infection with HIV, alcohol excess, haemolysis, Gilberts syndrome, thrombocytopaenia from haematological disease
Patient characteristics and setting	Centre details: single centre Sample size: 244 in total (80 in normal ALT value ('NALT') group; 164 in elevated ALT value ('EALT') group) Mean age: 46.35 (NALT group), 48.37 (EALT group) Gender (% male): 51.2 (NALT group), 57.3 (EALT group) Mean BMI: 23.9 (NALT group), 24.2 (EALT group) Mean ALT: 35.7 (NALT group), 134.9 (EALT group)
Index tests	Test name(s): Forns index Threshold(s) used: 4.2 and 6.9
Target condition and reference standard(s)	Target condition(s): F2+ Reference standard: liver biopsy Quality of liver biopsy: 15 mm in length and at least seven portal tracts
Flow and timing	Flow: all participants received the index test and none were excluded from the analysis Time between index test and biopsy: the same day as liver biopsy
Comparative	Comparators: APRI, FibroIndex, FibroTest, AST/ALT ratio
Notes	Reference Sebastiani 2008a represents the 'NALT' (normal ALT value) group Reference Sebastiani 2008b represents the 'EALT' (elevated ALT value) group

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Low risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index test (FIB-4)			
DOMAIN 2: Index test (Forns)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		

Sebastiani 2008a *(Continued)*

Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Low risk	

Sebastiani 2008b

Study characteristics	
Patient Sampling	Study design: cross-sectional diagnostic test accuracy study Method of sampling: cohort-based Direction of data collection: not reported Country: Italy Inclusion criteria: compensated chronic liver disease and diagnostic percutaneous liver biopsy Exclusion criteria: co-infection with HIV, alcohol excess, haemolysis, Gilberts syndrome, thrombocytopaenia from haematological disease
Patient characteristics and setting	Centre details: Single centre Sample size: 244 in total (80 in normal ALT value ('NALT') group; 164 in elevated ALT value ('EALT') group) Mean age: 46.35 (NALT group), 48.37 (EALT group) Gender (% male): 51.2 (NALT group), 57.3 (EALT group) Mean BMI: 23.9 (NALT group), 24.2 (EALT group) Mean ALT: 35.7 (NALT group), 134.9 (EALT group)
Index tests	Test name(s): Forns index Threshold(s) used: 4.2 and 6.9

Sebastiani 2008b *(Continued)*

Target condition and reference standard(s)	Target condition(s): F2+ Reference standard: liver biopsy Quality of liver biopsy: 15 mm in length and at least seven portal tracts in the specimen collected
Flow and timing	Flow: all participants received the index test and none were excluded from the analysis Time between index test and biopsy: the same day as liver biopsy
Comparative	Comparators: APRI, FibroIndex, FibroTest, AST/ALT ratio
Notes	Reference Sebastiani 2008a represents the 'NALT' (normal ALT value) group Reference Sebastiani 2008b represents the 'EALT' (elevated ALT value) group

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Low risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index test (FIB-4)			
DOMAIN 2: Index test (Forns)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		

Sebastiani 2008b *(Continued)*

Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Low risk	

Sebastiani 2012

Study characteristics	
Patient Sampling	Study design: cross-sectional diagnostic test accuracy study Method of sampling: cohort-based Direction of data collection: retrospective Country: Italy Inclusion criteria: compensated chronic HCV infection Exclusion criteria: co-infection with HIV, alcohol excess, haemolysis, Gilberts syndrome, thrombocytopaenia from haematological disease
Patient characteristics and setting	Centre details: multicentre: international study involving 7 centres accross Europe Sample size: 1013 Mean age: 48 Gender (% male): 56.7% Mean BMI: 24.6 Mean ALT: 102
Index tests	Test name(s): Forns index Threshold(s) used: 4.2 and 6.9 (for F2+)
Target condition and reference standard(s)	Target condition(s): F2+, F4 Reference standard: liver biopsy Quality of liver biopsy: a subgroup analysis was performed on participants with liver biopsies longer than 2 cm and containing more than 11 portal tracts. No information given about the length or portal tract numbers in the rest of the participants.
Flow and timing	Flow: all participants received the index test and none were excluded from the analysis Time between index test and biopsy: same day
Comparative	Comparators: FibroTest, APRI, SAFE biopsy algorithm, Fibropaca algorithm, Leroy algorithm

Sebastiani 2012 *(Continued)*

Notes

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Low risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index test (FIB-4)			
DOMAIN 2: Index test (Forns)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		

Sebastiani 2012 *(Continued)*

Could the patient flow have introduced bias?	Low risk

Segovia 2008

Study characteristics			
Patient Sampling	Study design: cross-sectional diagnostic test accuracy study Method of sampling: cohort-based Direction of data collection: retrospective Country: USA Inclusion criteria: patients after liver transplantation for HCV, infected with HCV, and had at least one biopsy report with FIB-4 calculated at the same time Exclusion criteria: not reported		
Patient characteristics and setting	Centre details: single centre Sample size: 219 Mean age: 52.3 Gender (% male): 68.95% Mean BMI: not reported Mean ALT: not reported Special characteristics: post liver transplantation		
Index tests	Test name(s): FIB-4 Threshold(s) used: 4.09		
Target condition and reference standard(s)	Target condition(s): F4 Reference standard: liver biopsy Quality of liver biopsy: not reported		
Flow and timing	Flow: all participants received the index test and none were excluded from the analysis Time between index test and biopsy: not reported		
Comparative	Comparators: nil		
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Unclear		
Could the selection of patients have introduced bias?		Unclear risk	
Are there concerns that the included patients and setting do not match the review question?			High

Segovia 2008 *(Continued)*

DOMAIN 2: Index test (FIB-4)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index test (Forns)			
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Unclear risk	

Shaikh 2009

Study characteristics	
Patient Sampling	Study design: cross-sectional diagnostic test accuracy study Method of sampling: cohort-based Direction of data collection: not reported Country: Hyderabad, India Inclusion criteria: anti-HCV and HCV RNA positive, untreated patients with chronic hepatitis C Exclusion criteria: HBV co-infection, regular alcohol intake, previous interferon treatment, and clinical or radiological evidence of cirrhosis (gastro-oesophageal varices, ascites, and hepatic encephalopathy)

Shaikh 2009 *(Continued)*

Patient characteristics and setting	Centre details: single centre Sample size: 158 Mean age: 36.7 Gender (% male): 69 Mean BMI: 26.1 Mean ALT: 77
Index tests	Test name(s): FIB-4 Threshold(s) used: 1.45
Target condition and reference standard(s)	Target condition(s): F3+ Reference standard: liver biopsy Quality of liver biopsy: greater than 10 mm and more than 5 portal tracts
Flow and timing	Flow: all participants received the index test and none were excluded from the analysis Time between index test and biopsy: 'before the biopsy' without further clarification
Comparative	Comparators: AST/ALT ratio, APRI
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Low risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index test (FIB-4)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index test (Forns)			

Shaikh 2009 *(Continued)*

DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Unclear risk	

Shiha 2017

Study characteristics	
Patient Sampling	Study design: cross-sectional diagnostic test accuracy study Method of sampling: cohort-based Direction of data collection: not reported Country: Egypt Inclusion criteria: anti-HCV and HCV RNA positive (HCV genotype 4) Exclusion criteria: people who had received any previous courses of antiviral or immunosuppressive therapy, HBV or HIV co-infection, any type of liver cancer
Patient characteristics and setting	Centre details: single centre Sample size: 604 Mean age: not reported Gender (% male): not reported Mean BMI: not reported Mean ALT: not reported Special characteristics: all participants had HCV genotype 4
Index tests	Test name(s): FIB-4 Threshold(s) used: 1.45
Target condition and reference standard(s)	Target condition(s): F2+ Reference standard: liver biopsy Quality of liver biopsy: 15 mm length, 4 portal tracts minimum
Flow and timing	Flow: all participants received the index test and none were excluded from the analysis

Shiha 2017 *(Continued)*

	Time between index test and biopsy: not reported		
Comparative	Comparators: FIB-5 (Fibrosis-5)		
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Low risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index test (FIB-4)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index test (Forns)			
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and timing			
Was there an appropriate interval between index test and reference standard?	Unclear		

Shiha 2017 *(Continued)*

Did all patients receive the same reference standard?	Yes	
Were all patients included in the analysis?	Yes	
Could the patient flow have introduced bias?		Unclear risk

Shiha 2022a

Study characteristics	
Patient Sampling	Study design: cross-sectional diagnostic test accuracy study Method of sampling: cohort-based Direction of data collection: retrospectively Country: Egypt, Japan, Kingdom of Saudi Arabia, Qatar, Turkey, Greece, Oman, and Jordan Inclusion criteria: adequate liver biopsy and complete biochemical and haematological data. All participants had active viraemia at the time of liver biopsy and clinical assessment, had compensated liver disease, and were treatment-naïve. Exclusion criteria: not reported
Patient characteristics and setting	Centre details: multicentre Sample size: 5417 Mean age: not reported Gender (% male): not reported Mean BMI: not reported Mean ALT: not reported
Index tests	Test name(s): FIB-4 Threshold(s) used: 1.45
Target condition and reference standard(s)	Target condition(s): F3+ Reference standard: liver biopsy Quality of liver biopsy: at least 20 mm core tissue or at least 11 portal tracts
Flow and timing	Flow: all participants received the index test but 106 were excluded from the analysis for incomplete blood tests. We deemed this to be an acceptable level, given the 5417 participants included in the analysis, and that there was an overall low risk of bias. Time between index test and biopsy: unclear
Comparative	Comparators: FIB-6, APRI, AAR
Notes	One main cohort from the authors. There were multiple international validation cohorts, meaning there was a total of 10 distinct cohorts available for data extraction. For the purposes of our meta-analysis, we extracted separate data for each cohort, as follows: • Main validation cohort (N = 5417) is under reference Shiha 2022a • External validation cohort from Menoufia (N = 400) is under reference Shiha 2022b • External validation cohort from Al-Azhar (N = 160) is under reference Shiha 2022c • External validation cohort from Japan (N = 797) is under reference Shiha 2022d • External validation cohort from KSA (N = 204) is under reference Shiha 2022e • External validation cohort from Qatar (N = 196) is under reference Shiha 2022f • External validation cohort from Turkey (N = 52) is under reference Shiha 2022g

Shiha 2022a *(Continued)*

- External validation cohort from Greece (N = 40) is under reference Shiha 2022h
- External validation cohort from Oman (N = 18) is under reference Shiha 2022i
- External validation cohort from Jordan (N = 10) is under reference Shiha 2022j

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Unclear risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index test (FIB-4)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index test (Forns)			
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern

Shiha 2022a *(Continued)*

DOMAIN 4: Flow and timing	
Was there an appropriate interval between index test and reference standard?	Unclear
Did all patients receive the same reference standard?	Yes
Were all patients included in the analysis?	Yes
Could the patient flow have introduced bias?	Unclear risk

Shiha 2022b

Study characteristics	
Patient Sampling	Study design: cross-sectional diagnostic test accuracy study Method of patient sampling: cohort-based Direction of data collection: retrospectively Country: Egypt, Japan, Kingdom of Saudi Arabia, Qatar, Turkey, Greece, Oman, and Jordan Inclusion criteria: adequate liver biopsy and complete biochemical and hematological data. All patients had active viremia at the time of liver biopsy and clinical assessment, had compensated liver disease, and were treatment naïve. Exclusion criteria: not reported
Patient characteristics and setting	Centre details: Multicentre Sample size: 5417 Mean age: not reported Gender (% male): not reported Mean BMI: not reported Mean ALT: not reported
Index tests	Test name(s): FIB-4 Threshold(s) used: 1.45
Target condition and reference standard(s)	Target condition(s): F3+ Reference standard: Liver biopsy Quality of liver biopsy: at least 20-mm core tissue or at least 11 portal tracts
Flow and timing	Flow: all patients received the index test but 106 were excluded from the analysis for incomplete blood tests. This is deemed to be an acceptable level compared to the 5417 patients included in analysis conferring overall low risk of bias. Time between index test and biopsy: unclear
Comparative	Comparators: FIB6, APRI, AAR
Notes	One main cohort from the authors. Multiple international validation cohorts meaning total number of cohorts available for data extraction was 10 distinct cohorts. Data were extracted on them separately for purposes of our meta-analysis. Main validation cohort (N = 5417) is under reference Shiha 2022a External validation cohort from Menoufia (N = 400) is under reference Shiha 2022b

Shiha 2022b *(Continued)*

	External validation cohort from Al-Azhar (N = 160) is under reference Shiha 2022c
	External validation cohort from Japan (N = 797) is under reference Shiha 2022d
	External validation cohort from KSA (N = 204) is under reference Shiha 2022e
	External validation cohort from Qatar (N = 196) is under reference Shiha 2022f
	External validation cohort from Turkey (N = 52) is under reference Shiha 2022g
	External validation cohort from Greece (N = 40) is under reference Shiha 2022h
	External validation cohort from Oman (N = 18) is under reference Shiha 2022i
	External validation cohort from Jordan (N = 10) is under reference Shiha 2022j

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Unclear risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index test (FIB-4)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index test (Forns)			
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	Yes		

Shiha 2022b *(Continued)*

Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Unclear risk	

Shiha 2022c

Study characteristics	
Patient Sampling	Study design: cross-sectional diagnostic test accuracy study Method of patient sampling: cohort-based Direction of data collection: retrospectively Country: Egypt, Japan, Kingdom of Saudi Arabia, Qatar, Turkey, Greece, Oman, and Jordan Inclusion criteria: adequate liver biopsy and complete biochemical and hematological data. All patients had active viremia at the time of liver biopsy and clinical assessment, had compensated liver disease, and were treatment naïve. Exclusion criteria: not reported
Patient characteristics and setting	Centre details: Multicentre Sample size: 5417 Mean age: not reported Gender (% male): not reported Mean BMI: not reported Mean ALT: not reported
Index tests	Test name(s): FIB-4 Threshold(s) used: 1.45
Target condition and reference standard(s)	Target condition(s): F3+ Reference standard: Liver biopsy Quality of liver biopsy: at least 20-mm core tissue or at least 11 portal tracts
Flow and timing	Flow: all patients received the index test but 106 were excluded from the analysis for incomplete blood tests. This is deemed to be an acceptable level compared to the 5417 patients included in analysis conferring overall low risk of bias. Time between index test and biopsy: unclear

Shiha 2022c *(Continued)*

Comparative	Comparators: FIB6, APRI, AAR
Notes	One main cohort from the authors. Multiple international validation cohorts meaning total number of cohorts available for data extraction was 10 distinct cohorts. Data were extracted on them separately for purposes of our meta-analysis. Main validation cohort (N = 5417) is under reference Shiha 2022a External validation cohort from Menoufia (N = 400) is under reference Shiha 2022b External validation cohort from Al-Azhar (N = 160) is under reference Shiha 2022c External validation cohort from Japan (N = 797) is under reference Shiha 2022d External validation cohort from KSA (N = 204) is under reference Shiha 2022e External validation cohort from Qatar (N = 196) is under reference Shiha 2022f External validation cohort from Turkey (N = 52) is under reference Shiha 2022g External validation cohort from Greece (N = 40) is under reference Shiha 2022h External validation cohort from Oman (N = 18) is under reference Shiha 2022i External validation cohort from Jordan (N = 10) is under reference Shiha 2022j

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Unclear risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index test (FIB-4)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	

Shiha 2022c *(Continued)*

Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index test (Forns)			
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Unclear risk	

Shiha 2022d

Study characteristics	
Patient Sampling	Study design: cross-sectional diagnostic test accuracy study Method of patient sampling: cohort-based Direction of data collection: retrospectively Country: Egypt, Japan, Kingdom of Saudi Arabia, Qatar, Turkey, Greece, Oman, and Jordan Inclusion criteria: adequate liver biopsy and complete biochemical and hematological data. All patients had active viremia at the time of liver biopsy and clinical assessment, had compensated liver disease, and were treatment naïve. Exclusion criteria: not reported
Patient characteristics and setting	Centre details: Multicentre Sample size: 5417 Mean age: not reported Gender (% male): not reported Mean BMI: not reported Mean ALT: not reported
Index tests	Test name(s): FIB-4

Shiha 2022d *(Continued)*

	Threshold(s) used: 1.45
Target condition and reference standard(s)	Target condition(s): F3+ Reference standard: Liver biopsy Quality of liver biopsy: at least 20-mm core tissue or at least 11 portal tracts
Flow and timing	Flow: all patients received the index test but 106 were excluded from the analysis for incomplete blood tests. This is deemed to be an acceptable level compared to the 5417 patients included in analysis conferring overall low risk of bias. Time between index test and biopsy: unclear
Comparative	Comparators: FIB6, APRI, AAR
Notes	One main cohort from the authors. Multiple international validation cohorts meaning total number of cohorts available for data extraction was 10 distinct cohorts. Data were extracted on them separately for purposes of our meta-analysis. Main validation cohort (N = 5417) is under reference Shiha 2022a External validation cohort from Menoufia (N = 400) is under reference Shiha 2022b External validation cohort from Al-Azhar (N = 160) is under reference Shiha 2022c External validation cohort from Japan (N = 797) is under reference Shiha 2022d External validation cohort from KSA (N = 204) is under reference Shiha 2022e External validation cohort from Qatar (N = 196) is under reference Shiha 2022f External validation cohort from Turkey (N = 52) is under reference Shiha 2022g External validation cohort from Greece (N = 40) is under reference Shiha 2022h External validation cohort from Oman (N = 18) is under reference Shiha 2022i External validation cohort from Jordan (N = 10) is under reference Shiha 2022j

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Unclear risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index test (FIB-4)			

Shiha 2022d *(Continued)*

Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index test (Forns)			
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Unclear risk	

Shiha 2022e

Study characteristics	
Patient Sampling	Study design: cross-sectional diagnostic test accuracy study Method of patient sampling: cohort-based Direction of data collection: retrospectively Country: Egypt, Japan, Kingdom of Saudi Arabia, Qatar, Turkey, Greece, Oman, and Jordan Inclusion criteria: adequate liver biopsy and complete biochemical and hematological data. All patients had active viremia at the time of liver biopsy and clinical assessment, had compensated liver disease, and were treatment naïve.

Shiha 2022e *(Continued)*

	Exclusion criteria: not reported
Patient characteristics and setting	Centre details: Multicentre Sample size: 5417 Mean age: not reported Gender (% male): not reported Mean BMI: not reported Mean ALT: not reported
Index tests	Test name(s): FIB-4 Threshold(s) used: 1.45
Target condition and reference standard(s)	Target condition(s): F3+ Reference standard: Liver biopsy Quality of liver biopsy: at least 20-mm core tissue or at least 11 portal tracts
Flow and timing	Flow: all patients received the index test but 106 were excluded from the analysis for incomplete blood tests. This is deemed to be an acceptable level compared to the 5417 patients included in analysis conferring overall low risk of bias. Time between index test and biopsy: unclear
Comparative	Comparators: FIB6, APRI, AAR
Notes	One main cohort from the authors. Multiple international validation cohorts meaning total number of cohorts available for data extraction was 10 distinct cohorts. Data were extracted on them separately for purposes of our meta-analysis. Main validation cohort (N = 5417) is under reference Shiha 2022a External validation cohort from Menoufia (N = 400) is under reference Shiha 2022b External validation cohort from Al-Azhar (N = 160) is under reference Shiha 2022c External validation cohort from Japan (N = 797) is under reference Shiha 2022d External validation cohort from KSA (N = 204) is under reference Shiha 2022e External validation cohort from Qatar (N = 196) is under reference Shiha 2022f External validation cohort from Turkey (N = 52) is under reference Shiha 2022g External validation cohort from Greece (N = 40) is under reference Shiha 2022h External validation cohort from Oman (N = 18) is under reference Shiha 2022i External validation cohort from Jordan (N = 10) is under reference Shiha 2022j

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	Yes		

Shiha 2022e *(Continued)*

Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Unclear risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index test (FIB-4)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index test (Forns)			
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Unclear risk	

Shiha 2022f

Study characteristics	
Patient Sampling	Study design: cross-sectional diagnostic test accuracy study Method of patient sampling: cohort-based Direction of data collection: retrospectively Country: Egypt, Japan, Kingdom of Saudi Arabia, Qatar, Turkey, Greece, Oman, and Jordan Inclusion criteria: adequate liver biopsy and complete biochemical and hematological data. All patients had active viremia at the time of liver biopsy and clinical assessment, had compensated liver disease, and were treatment naïve. Exclusion criteria: not reported
Patient characteristics and setting	Centre details: Multicentre Sample size: 5417 Mean age: not reported Gender (% male): not reported Mean BMI: not reported Mean ALT: not reported
Index tests	Test name(s): FIB4 Threshold(s) used: 1.45
Target condition and reference standard(s)	Target condition(s): F3+ Reference standard: Liver biopsy Quality of liver biopsy: at least 20-mm core tissue or at least 11 portal tracts
Flow and timing	Flow: all patients received the index test but 106 were excluded from the analysis for incomplete blood tests. This is deemed to be an acceptable level compared to the 5417 patients included in analysis conferring overall low risk of bias. Time between index test and biopsy: unclear
Comparative	Comparators: FIB6, APRI, AAR
Notes	One main cohort from the authors. Multiple international validation cohorts meaning total number of cohorts available for data extraction was 10 distinct cohorts. Data were extracted on them separately for purposes of our meta-analysis. Main validation cohort (N = 5417) is under reference Shiha 2022a External validation cohort from Menoufia (N = 400) is under reference Shiha 2022b External validation cohort from Al-Azhar (N = 160) is under reference Shiha 2022c External validation cohort from Japan (N = 797) is under reference Shiha 2022d External validation cohort from KSA (N = 204) is under reference Shiha 2022e External validation cohort from Qatar (N = 196) is under reference Shiha 2022f External validation cohort from Turkey (N = 52) is under reference Shiha 2022g External validation cohort from Greece (N = 40) is under reference Shiha 2022h External validation cohort from Oman (N = 18) is under reference Shiha 2022i External validation cohort from Jordan (N = 10) is under reference Shiha 2022j
Methodological quality	

Shiha 2022f *(Continued)*

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Unclear risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index test (FIB-4)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index test (Forns)			
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and timing			
Was there an appropriate interval between index test and reference standard?	Unclear		

Shiha 2022f *(Continued)*

Did all patients receive the same reference standard?	Yes
Were all patients included in the analysis?	Yes
Could the patient flow have introduced bias?	Unclear risk

Shiha 2022g

Study characteristics	
Patient Sampling	Study design: cross-sectional diagnostic test accuracy study Method of patient sampling: cohort-based Direction of data collection: retrospectively Country: Egypt, Japan, Kingdom of Saudi Arabia, Qatar, Turkey, Greece, Oman, and Jordan Inclusion criteria: adequate liver biopsy and complete biochemical and hematological data. All patients had active viremia at the time of liver biopsy and clinical assessment, had compensated liver disease, and were treatment naïve. Exclusion criteria: not reported
Patient characteristics and setting	Centre details: Multicentre Sample size: 5417 Mean age: not reported Gender (% male): not reported Mean BMI: not reported Mean ALT: not reported
Index tests	Test name(s): FIB-4 Threshold(s) used: 1.45
Target condition and reference standard(s)	Target condition(s): F3+ Reference standard: Liver biopsy Quality of liver biopsy: at least 20-mm core tissue or at least 11 portal tracts
Flow and timing	Flow: all patients received the index test but 106 were excluded from the analysis for incomplete blood tests. This is deemed to be an acceptable level compared to the 5417 patients included in analysis conferring overall low risk of bias. Time between index test and biopsy: unclear
Comparative	Comparators: FIB6, APRI, AAR
Notes	One main cohort from the authors. Multiple international validation cohorts meaning total number of cohorts available for data extraction was 10 distinct cohorts. Data were extracted on them separately for purposes of our meta-analysis. Main validation cohort (N = 5417) is under reference Shiha 2022a External validation cohort from Menoufia (N = 400) is under reference Shiha 2022b External validation cohort from Al-Azhar (N = 160) is under reference Shiha 2022c External validation cohort from Japan (N = 797) is under reference Shiha 2022d External validation cohort from KSA (N = 204) is under reference Shiha 2022e

Shiha 2022g *(Continued)*

External validation cohort from Qatar (N = 196) is under reference Shiha 2022f

External validation cohort from Turkey (N = 52) is under reference Shiha 2022g

External validation cohort from Greece (N = 40) is under reference Shiha 2022h

External validation cohort from Oman (N = 18) is under reference Shiha 2022i

External validation cohort from Jordan (N = 10) is under reference Shiha 2022j

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Unclear risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index test (FIB-4)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index test (Forns)			
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	

Shiha 2022g *(Continued)*

Are there concerns that the target condition as defined by the reference standard does not match the question?		Low concern
DOMAIN 4: Flow and timing		
Was there an appropriate interval between index test and reference standard?	Unclear	
Did all patients receive the same reference standard?	Yes	
Were all patients included in the analysis?	Yes	
Could the patient flow have introduced bias?		Unclear risk

Shiha 2022h

Study characteristics	
Patient Sampling	Study design: cross-sectional diagnostic test accuracy study Method of patient sampling: cohort-based Direction of data collection: retrospectively Country: Egypt, Japan, Kingdom of Saudi Arabia, Qatar, Turkey, Greece, Oman, and Jordan Inclusion criteria: adequate liver biopsy and complete biochemical and hematological data. All patients had active viremia at the time of liver biopsy and clinical assessment, had compensated liver disease, and were treatment naïve. Exclusion criteria: not reported
Patient characteristics and setting	Centre details: Multicentre Sample size: 5417 Mean age: not reported Gender (% male): not reported Mean BMI: not reported Mean ALT: not reported
Index tests	Test name(s): FIB-4 Threshold(s) used: 1.45
Target condition and reference standard(s)	Target condition(s): F3+ Reference standard: Liver biopsy Quality of liver biopsy: at least 20-mm core tissue or at least 11 portal tracts
Flow and timing	Flow: all patients received the index test but 106 were excluded from the analysis for incomplete blood tests. This is deemed to be an acceptable level compared to the 5417 patients included in analysis conferring overall low risk of bias. Time between index test and biopsy: unclear
Comparative	Comparators: FIB6, APRI, AAR
Notes	One main cohort from the authors. Multiple international validation cohorts meaning total number of cohorts available for data extraction was 10 distinct cohorts. Data were extracted on them separately for purposes of our meta-analysis.

Shiha 2022h *(Continued)*

	Main validation cohort (N = 5417) is under reference Shiha 2022a
	External validation cohort from Menoufia (N = 400) is under reference Shiha 2022b
	External validation cohort from Al-Azhar (N = 160) is under reference Shiha 2022c
	External validation cohort from Japan (N = 797) is under reference Shiha 2022d
	External validation cohort from KSA (N = 204) is under reference Shiha 2022e
	External validation cohort from Qatar (N = 196) is under reference Shiha 2022f
	External validation cohort from Turkey (N = 52) is under reference Shiha 2022g
	External validation cohort from Greece (N = 40) is under reference Shiha 2022h
	External validation cohort from Oman (N = 18) is under reference Shiha 2022i
	External validation cohort from Jordan (N = 10) is under reference Shiha 2022j

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Unclear risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index test (FIB-4)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index test (Forns)			
DOMAIN 3: Reference standard			

Shiha 2022h *(Continued)*

Is the reference standard likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Unclear risk	

Shiha 2022i

Study characteristics	
Patient Sampling	Study design: cross-sectional diagnostic test accuracy study Method of patient sampling: cohort-based Direction of data collection: retrospectively Country: Egypt, Japan, Kingdom of Saudi Arabia, Qatar, Turkey, Greece, Oman, and Jordan Inclusion criteria: adequate liver biopsy and complete biochemical and hematological data. All patients had active viremia at the time of liver biopsy and clinical assessment, had compensated liver disease, and were treatment naïve. Exclusion criteria: not reported
Patient characteristics and setting	Centre details: Multicentre Sample size: 5417 Mean age: not reported Gender (% male): not reported Mean BMI: not reported Mean ALT: not reported
Index tests	Test name(s): FIB-4 Threshold(s) used: 1.45
Target condition and reference standard(s)	Target condition(s): F3+ Reference standard: Liver biopsy Quality of liver biopsy: at least 20-mm core tissue or at least 11 portal tracts

Shiha 2022i *(Continued)*

Flow and timing	Flow: all patients received the index test but 106 were excluded from the analysis for incomplete blood tests. This is deemed to be an acceptable level compared to the 5417 patients included in analysis conferring overall low risk of bias. Time between index test and biopsy: unclear
Comparative	Comparators: FIB6, APRI, AAR
Notes	One main cohort from the authors. Multiple international validation cohorts meaning total number of cohorts available for data extraction was 10 distinct cohorts. Data were extracted on them separately for purposes of our meta-analysis. Main validation cohort (N = 5417) is under reference Shiha 2022a External validation cohort from Menoufia (N = 400) is under reference Shiha 2022b External validation cohort from Al-Azhar (N = 160) is under reference Shiha 2022c External validation cohort from Japan (N = 797) is under reference Shiha 2022d External validation cohort from KSA (N = 204) is under reference Shiha 2022e External validation cohort from Qatar (N = 196) is under reference Shiha 2022f External validation cohort from Turkey (N = 52) is under reference Shiha 2022g External validation cohort from Greece (N = 40) is under reference Shiha 2022h External validation cohort from Oman (N = 18) is under reference Shiha 2022i External validation cohort from Jordan (N = 10) is under reference Shiha 2022j

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Unclear risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index test (FIB-4)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		

Shiha 2022i *(Continued)*

If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index test (Forns)			
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Unclear risk	

Shiha 2022j

Study characteristics	
Patient Sampling	Study design: cross-sectional diagnostic test accuracy study Method of patient sampling: cohort-based Direction of data collection: retrospectively Country: Egypt, Japan, Kingdom of Saudi Arabia, Qatar, Turkey, Greece, Oman, and Jordan Inclusion criteria: adequate liver biopsy and complete biochemical and hematological data. All patients had active viremia at the time of liver biopsy and clinical assessment, had compensated liver disease, and were treatment naïve. Exclusion criteria: not reported
Patient characteristics and setting	Centre details: Multicentre Sample size: 5417 Mean age: not reported

Shiha 2022j *(Continued)*

	Gender (% male): not reported Mean BMI: not reported Mean ALT: not reported
Index tests	Test name(s): FIB-4 Threshold(s) used: 1.45
Target condition and reference standard(s)	Target condition(s): F3+ Reference standard: Liver biopsy Quality of liver biopsy: at least 20-mm core tissue or at least 11 portal tracts
Flow and timing	Flow: all patients received the index test but 106 were excluded from the analysis for incomplete blood tests. This is deemed to be an acceptable level compared to the 5417 patients included in analysis conferring overall low risk of bias. Time between index test and biopsy: unclear
Comparative	Comparators: FIB6, APRI, AAR
Notes	One main cohort from the authors. Multiple international validation cohorts meaning total number of cohorts available for data extraction was 10 distinct cohorts. Data were extracted on them separately for purposes of our meta-analysis. Main validation cohort (N = 5417) is under reference Shiha 2022a External validation cohort from Menoufia (N = 400) is under reference Shiha 2022b External validation cohort from Al-Azhar (N = 160) is under reference Shiha 2022c External validation cohort from Japan (N = 797) is under reference Shiha 2022d External validation cohort from KSA (N = 204) is under reference Shiha 2022e External validation cohort from Qatar (N = 196) is under reference Shiha 2022f External validation cohort from Turkey (N = 52) is under reference Shiha 2022g External validation cohort from Greece (N = 40) is under reference Shiha 2022h External validation cohort from Oman (N = 18) is under reference Shiha 2022i External validation cohort from Jordan (N = 10) is under reference Shiha 2022j

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Unclear risk	

Shiha 2022j *(Continued)*

Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index test (FIB-4)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index test (Forns)			
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Unclear risk	

Silva Junior 2014

Study characteristics	
Patient Sampling	Study design: cross-sectional diagnostic test accuracy study Method of sampling: cohort-based

Silva Junior 2014 *(Continued)*

	Direction of data collection: prospective, consecutive Country: Brazil Inclusion criteria: chronic hepatitis C Exclusion criteria: HIV co-infection, hepatitis B virus co-infection, chronic alcohol abuse, cholestatic chronic hepatitis, non-alcoholic steatohepatitis, autoimmune chronic hepatitis, haemochromatosis, Wilson's disease, hepatocellular carcinoma, prior liver transplantation, prior interferon therapy, immunosuppressive therapy, insufficient liver tissue for staging of fibrosis
Patient characteristics and setting	Centre details: single centre Sample size: 51 Mean age: 53.8 Gender (% male): 35.3% Mean BMI: 25.16 Mean ALT: 60.55
Index tests	Test name(s): FIB-4 and Forns index Threshold(s) used: FIB-4 (3.25) and Forns index (5.73)
Target condition and reference standard(s)	Target condition(s): FIB-4 F3+, Forns index F2+ Reference standard: liver biopsy Quality of liver biopsy: at least 15 mm length but number of portal tracts not reported
Flow and timing	Flow: all participants received the index test and none were excluded from the analysis Time between index test and biopsy: maximum 6 months
Comparative	Comparators: APRI, King's, ARFI
Notes	

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Low risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index test (FIB-4)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		

Silva Junior 2014 *(Continued)*

Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index test (Forns)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and timing			
Was there an appropriate interval between index test and reference standard?	No		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Stauber 2015

Study characteristics	
Patient Sampling	Study design: cross-sectional diagnostic test accuracy study Method of sampling: cohort-based Direction of data collection: not reported Country: Austria Inclusion criteria: chronic hepatitis C Exclusion criteria: prior antiviral treatment, HCC, previous liver transplantation, lack of a representative biopsy specimen

Stauber 2015 *(Continued)*

Patient characteristics and setting	Centre details: two centres Sample size: 614 Mean age: not reported Gender (% male): not reported Mean BMI: not reported Mean ALT: not reported
Index tests	Test name(s): FIB-4 Threshold(s) used: 1.77
Target condition and reference standard(s)	Target condition(s): F3+ Reference standard: liver biopsy Quality of liver biopsy: not reported
Flow and timing	Flow: all participants received the index test and none were excluded from the analysis Time between index test and biopsy: at time of biopsy
Comparative	Comparators: APRI, platelets, Lok, LSM
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Low risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index test (FIB-4)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index test (Forns)			
DOMAIN 3: Reference standard			

Stauber 2015 *(Continued)*

Is the reference standard likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Low risk	

Sterling 2006

Study characteristics	
Patient Sampling	Study design: cross-sectional diagnostic test accuracy study Method of sampling: cohort-based Direction of data collection: retrospective Country: USA, Canada, Spain, Belgium, Germany, UK, Brazil Inclusion criteria: "age>18 years, infection with both HIV and HCV, elevated ALT levels on 2 or more occasions within the previous 12 months with compensated liver disease, liver biopsy with liver histology consistent with chronic HCV. HIV antibodies +, detectable HCV RNA, and HCV treatment–naïve. Patients with CD4 counts of 100-199 cells/mm3 were eligible if the HIV RNA load was <5,000 copies/mL. Stable HAART regimen for at least 6 weeks prior to study entry or not on any antiretroviral treatment at least 8 weeks prior to randomization." Exclusion criteria: "active HIV-related opportunistic infection or cancer; an absolute neutrophil count <1,500 cells/mm3; platelet<70,000/mm3; haemoglobin<11 g/dL for women and <12 g/dL for men; creatinine level >1.5 times the upper limit of normal; concurrent hepatitis A or B infection; evidence of decompensated liver disease; severe psychiatric disease; clinically significant coexisting medical conditions that would preclude HCV therapy; or previous treatment with interferon or ribavirin."
Patient characteristics and setting	Centre details: multicentre Sample size: 277 (validation set) Mean age: 40 Gender (% male): 82% Mean BMI: 24 Mean ALT: 85 Special characteristics: all HIV co-infected patients
Index tests	Test name(s): FIB-4 Threshold(s) used: 1.45 and 3.25 for F3+; 0.6 and 1 for F2+

Sterling 2006 *(Continued)*

Target condition and reference standard(s)	Target condition(s): F2+ and F3+ Reference standard: liver biopsy Quality of liver biopsy: not reported
Flow and timing	Flow: 36 participants were excluded because of lack of interpretable histology Time between index test and biopsy: average 98 days
Comparative	Comparators: nil
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Low risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index test (FIB-4)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index test (Forns)			
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	

Sterling 2006 *(Continued)*

Are there concerns that the target condition as defined by the reference standard does not match the question?		Low concern
DOMAIN 4: Flow and timing		
Was there an appropriate interval between index test and reference standard?	No	
Did all patients receive the same reference standard?	Yes	
Were all patients included in the analysis?	No	
Could the patient flow have introduced bias?		High risk

Stibbe 2011

Study characteristics	
Patient Sampling	Study design: cross-sectional diagnostic test accuracy study Method of sampling: cohort-based Direction of data collection: not reported Country: the Netherlands Inclusion criteria: mono-infected people with chronic hepatitis B or chronic hepatitis C referred for liver biopsy Exclusion criteria: alcohol intake > 20 g/day, co-infection with HIV or hepatitis D, or the presence of hepatocellular carcinoma. Additionally, healthy controls were included for the breath tests and serological tests.
Patient characteristics and setting	Centre details: single centre Sample size: 41 (people with chronic hepatitis C) Mean age: 47 Gender (% male): 65.8% Mean BMI: 25 Mean ALT: not reported
Index tests	Test name(s): FIB-4 Threshold(s) used: 1.45, 3.25
Target condition and reference standard(s)	Target condition(s): F3+ Reference standard: liver biopsy Quality of liver biopsy: minimum length 20 mm but number of portal tracts not reported
Flow and timing	Flow: all participants received the index test and none were excluded from the analysis Time between index test and biopsy: same day
Comparative	Comparators: breath tests, hyaluronic acid (HA), APRI, FibroTest, transient elastography
Notes	

Stibbe 2011 *(Continued)*

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Unclear risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index test (FIB-4)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index test (Forns)			
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Low risk	

Sène 2006

Study characteristics	
Patient Sampling	Study design: cross-sectional diagnostic test accuracy study Method of sampling: cohort-based Direction of data collection: not reported Country: France Inclusion criteria: people infected with HCV, with a liver biopsy report and with Forns index calculated with blood samples obtained within 2 months from biopsy Exclusion criteria: not reported
Patient characteristics and setting	Centre details: single centre Sample size: 138 Mean age: 57.9 Gender (% male): 50 Mean BMI: not reported Mean ALT: not reported Special characteristics: 72 of 138 had cryoglobulinaemia
Index tests	Test name(s): Forns index Threshold(s) used: 9.29
Target condition and reference standard(s)	Target condition(s): F2+ Reference standard: liver biopsy Quality of liver biopsy: a good quality biopsy (5+ portal tracts) was obtained in 43 participants (31.2%)
Flow and timing	Flow: all participants received the index test and none were excluded from the analysis Time between index test and biopsy: 92.8% of participants received the index test within 2 months of the biopsy, but the timing for the remaining participants is unclear.
Comparative	Comparators: APRI, Fibrotest–Actitest (FT–AT), Age–platelet index, platelets, hyaluronic acid (HA)
Notes	

Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Low risk	
Are there concerns that the included patients and setting do not match the review question?			High

Sène 2006 *(Continued)*

DOMAIN 2: Index test (FIB-4)			
DOMAIN 2: Index test (Forns)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Unclear risk	

Tachi 2015

Study characteristics	
Patient Sampling	Study design: cross-sectional diagnostic test accuracy study Method of sampling: cohort-based Direction of data collection: not reported Country: Taiwan Inclusion criteria: people with HCV who received antiviral therapy (data reported separately pre-treatment) Exclusion criteria: antibodies against human immunodeficiency virus or hepatitis B virus surface antigen, excessive active alcohol consumption (daily intake > 40 g of ethanol) or drug abuse, or other forms of liver disease (e.g. autoimmune hepatitis, alcoholic liver

Tachi 2015 *(Continued)*

	disease, or haemochromatosis), previous history of hepatocellular carcinoma
Patient characteristics and setting	Centre details: single centre Sample size: 115 (before treatment cohort) Mean age: 64 Gender (% male): 60.9 Mean BMI: not reported Mean ALT: 19
Index tests	Test name(s): FIB-4, Forns index Threshold(s) used: FIB-4 2.84 (F2+) and 3.97 (F3+), Forns index 7.56 (F2+ and F3+)
Target condition and reference standard(s)	Target condition(s): F2+ and F3+ Reference standard: liver biopsy Quality of liver biopsy: median biopsy length 28 mm. No info about minimum requirements nor number of portal tracts.
Flow and timing	Flow: although there were exclusions for the post-treatment group, all participants who received the index test and biopsy were included in the analysis of the pre-treatment group Time between index test and biopsy: same day
Comparative	Comparators: APRI
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Low risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index test (FIB-4)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern

Tachi 2015 *(Continued)*

DOMAIN 2: Index test (Forns)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Low risk	

Tanwar 2017

Study characteristics	
Patient Sampling	Study design: cross-sectional diagnostic test accuracy study Method of sampling: cohort-based Direction of data collection: prospective Country: Germany and Austria Inclusion criteria: from 18 to 65 years old with evidence of CHC Exclusion criteria: acute hepatitis, therapy with steroids or immunosuppressive drugs in the previous 3 months, Child-Pugh stage B or C cirrhosis, thrombocytopenia, other chronic liver diseases, autoimmune diseases, HIV infection, alcohol abuse, active drug abuse, pregnancy, or psychiatric diseases including depression
Patient characteristics and setting	Centre details: multicentre (18 centres in Germany and Austria) Sample size: 80

Tanwar 2017 *(Continued)*

	Mean age: 48.9 Gender (% male): 57 Mean BMI: 24.3 Mean ALT: 81.6 Special characteristics: all participants were non-responders to standard therapy with either interferon or pegylated interferon and ribavirin
Index tests	Test name(s): FIB-4, Forns index Threshold(s) used: FIB-4 1 (F2+) and 1.45 (F3+), Forns index 5 (F2+)
Target condition and reference standard(s)	Target condition(s): F2+ and F3+ Reference standard: liver biopsy Quality of liver biopsy: not reported
Flow and timing	Flow: all participants received the index test and none were excluded from the analysis Time between index test and biopsy: up to 6 months
Comparative	Comparators: AST to ALT ratio, AST to Platelet Ratio Index (APRI), Fibrometer, HepaScore, ELF, Fibrospect II, and HA
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Low risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index test (FIB-4)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index test (Forns)			

Tanwar 2017 *(Continued)*

Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and timing			
Was there an appropriate interval between index test and reference standard?	No		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Toson 2017

Study characteristics	
Patient Sampling	Study design: cross-sectional diagnostic test accuracy study Method of sampling: cohort-based Direction of data collection: prospective, consecutive Country: Egypt Inclusion criteria: people with CHC Exclusion criteria: HAV, HBV, age > 70, alcohol excess, HCC, previous interferon treatment, decompensated liver disease (ascites, jaundice, variceal haemorrhage, or encephalopathy), evidence of coexistent liver disease, and liver transplantation
Patient characteristics and setting	Centre details: single centre Sample size: 52 (validation group) Mean age: 46.1 Gender (% male): 67 Mean BMI: not reported

Toson 2017 *(Continued)*

	Mean ALT: 59.1
Index tests	Test name(s): FIB-4 Threshold(s) used: not interpretable
Target condition and reference standard(s)	Target condition(s): F2+, F3+, F4 Reference standard: liver biopsy Quality of liver biopsy: minimum 15 mm or containing five portal tracts, except for cirrhosis, for which no limitation was required
Flow and timing	Flow: all participants received the index test and none were excluded from the analysis Time between index test and biopsy: maximum 2 weeks
Comparative	Comparators: angiogenic markers including hepatocyte growth factor (HGF), basic fibroblast growth factor (b-FGF), angiopoietin-2, and endostatin (ES).
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Low risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index test (FIB-4)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index test (Forns)			
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	No		

Toson 2017 *(Continued)*

Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Low risk	

Trang 2008

Study characteristics	
Patient Sampling	Study design: cross-sectional diagnostic test accuracy study Method of sampling: cohort-based Direction of data collection: retrospective Country: USA Inclusion criteria: people with HIV + HCV Exclusion criteria: HBV, missing laboratory values or HCC
Patient characteristics and setting	Centre details: single centre Sample size: 81 Mean age: 46.7 Gender (% male): 68 Mean BMI: not reported Mean ALT: 104 Special characteristics: HIV co-infected patients
Index tests	Test name(s): FIB-4 Threshold(s) used: 1,39-1.86-2.05 for F2+; 1.45-2.22-3.25 for F3+
Target condition and reference standard(s)	Target condition(s): F2+ and F3+ Reference standard: liver biopsy Quality of liver biopsy: not reported
Flow and timing	Flow: all participants received the index test and none were excluded from the analysis Time between index test and biopsy: within 6 months
Comparative	Comparators: APRI
Notes	
Methodological quality	

Trang 2008 *(Continued)*

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Low risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index test (FIB-4)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index test (Forns)			
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and timing			
Was there an appropriate interval between index test and reference standard?	No		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Trifan 2009

Study characteristics			
Patient Sampling	Study design: cross-sectional diagnostic test accuracy study Method of sampling: cohort-based Direction of data collection: not reported Country: Romania Inclusion criteria: HCV mono-infected and treatment-naive patients Exclusion criteria: not reported		
Patient characteristics and setting	Centre details: multicentre (5 centres) Sample size: 312 Mean age: 47 Gender (% male): 32.2 Mean BMI: not reported Mean ALT: not reported		
Index tests	Test name(s): FIB-4, Forns index Threshold(s) used: Forns index 5.35 (F2+) and 6.38 (F3+), FIB-4 1.66 (F2+) and 1.36 (F3+)		
Target condition and reference standard(s)	Target condition(s): F2+ and F3+ Reference standard: liver biopsy Quality of liver biopsy: more than 6 portal tracts		
Flow and timing	Flow: all participants received the index test and none were excluded from the analysis Time between index test and biopsy: 48 hours		
Comparative	Comparators: APRI, Forns index, FibroScan, Hitachi EUB-8500		
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Low risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index test (FIB-4)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		

Trifan 2009 *(Continued)*

If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index test (Forns)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Low risk	

Tsukano 2017

Study characteristics	
Patient Sampling	Study design: cross-sectional diagnostic test accuracy study Method of sampling: cohort-based Direction of data collection: not reported Country: Japan Inclusion criteria: HCV+ patients

Tsukano 2017 *(Continued)*

	Exclusion criteria: people with HBV co-infection or other various liver diseases, or who were 18 years of age or younger
Patient characteristics and setting	Centre details: single centre Sample size: 302 Mean age: 59 Gender (% male): 49 Mean BMI: 23.3 Mean ALT: 66
Index tests	Test name(s): FIB-4 Threshold(s) used: 0.805, 0.816, 0.887
Target condition and reference standard(s)	Target condition(s): F2+, F3+, F4 Reference standard: liver biopsy Quality of liver biopsy: at least 17 mm in length and containing 5 to 8 portal regions
Flow and timing	Flow: all participants received the index test and none were excluded from the analysis Time between index test and biopsy: 1 week
Comparative	Comparators: APRI, VTQ (virtual touch quantification)
Notes	

Methodological quality

Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Low risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index test (FIB-4)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index test (Forns)			

Tsukano 2017 *(Continued)*

DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	No		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		High risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Low risk	

Tural 2009

Study characteristics	
Patient Sampling	Study design: cross-sectional diagnostic test accuracy study Method of sampling: cohort-based Direction of data collection: not reported Country: Spain Inclusion criteria: a positive HCV RNA determination in serum, alanine transaminase (ALT) levels persistently above the upper limit of normal, and being naive for HCV antiviral therapy Exclusion criteria: alcohol excess, active intravenous drug use, HBV, decompensated cirrhosis, uncontrolled psychiatric illness, active infection or cancer, high neutrophils or low Hb (haemoglobin)
Patient characteristics and setting	Centre details: single centre Sample size: 324 Mean age: 38 (median) Gender (% male): 72 Mean BMI: not reported Mean ALT: 89 Special characteristics: HIV and HCV co-infection
Index tests	Test name(s): FIB-4, Forns index Threshold(s) used: Forns index: 4.20, 5.51, 6.90; Fib-4: 1.45, 3.25
Target condition and reference standard(s)	Target condition(s): F2+, F3+ Reference standard: liver biopsy Quality of liver biopsy: not reported

Tural 2009 *(Continued)*

Flow and timing	Flow: all participants received the index test and none were excluded from the analysis Time between index test and biopsy: maximum 120 days
Comparative	Comparators: APRI
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Low risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index test (FIB-4)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index test (Forns)			
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and timing			

Tural 2009 *(Continued)*

Was there an appropriate interval between index test and reference standard?	No		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		High risk	

Udompap 2020

Study characteristics			
Patient Sampling	Study design: cross-sectional diagnostic test accuracy study Method of sampling: cohort-based Direction of data collection: prospective Country: USA Inclusion criteria: either chronic HBV or HCV mono-infection Exclusion criteria: people with HBV and HCV co-infection or HIV infection or those who received any antiviral therapy		
Patient characteristics and setting	Centre details: single centre Sample size: 60 Mean age: 51 Gender (% male): 43 Mean BMI: not reported Mean ALT: 74		
Index tests	Test name(s): FIB-4 Threshold(s) used: 1.29, 1.56, 2.59		
Target condition and reference standard(s)	Target condition(s): F2+, F3+, F4 Reference standard: liver biopsy Quality of liver biopsy: not reported		
Flow and timing	Flow: all participants received the index test and none were excluded from the analysis Time between index test and biopsy: not reported		
Comparative	Comparators: APRI, SWE (shear wave elastography), FibroScan		
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		

Udompap 2020 *(Continued)*

Could the selection of patients have introduced bias?		Unclear risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index test (FIB-4)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index test (Forns)			
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Unclear risk	

Usluer 2012

Study characteristics	
Patient Sampling	Study design: cross-sectional diagnostic test accuracy study Method of sampling: cohort-based Direction of data collection: prospective Country: Turkey Inclusion criteria: anti-HCV and HCV-RNA positive, genotype 1, treatment-naive, CHC patients between 18 and 65 years of age

Usluer 2012 *(Continued)*

	Exclusion criteria: concomitant chronic liver disease, decompensated cirrhosis, HCC, alcohol or drug abuse, HBV, HDV, HIV, use of medications which may cause haemolysis, acute hepatitis and inflammation, extrahepatic cholestasis
Patient characteristics and setting	Centre details: multicentre (14 centre) Sample size: 77 Mean age: 49 Gender (% male): 43 Mean BMI: not reported Mean ALT: 63 Special characteristics: all participants had genotype 1 HCV
Index tests	Test name(s): FIB-4 Threshold(s) used: 1.45
Target condition and reference standard(s)	Target condition(s): F2+ Reference standard: liver biopsy Quality of liver biopsy: minimum 2 cm in length, but number of portal tracts not reported.
Flow and timing	Flow: all participants received the index test and none were excluded from the analysis Time between index test and biopsy: same day
Comparative	Comparators: APRI, FibroTest, AST to platelet ration, Actitest, CDS, platelets, AAR, GUCI
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Low risk	
Are there concerns that the included patients and setting do not match the review question?			High
DOMAIN 2: Index test (FIB-4)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	

Usluer 2012 *(Continued)*

Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index test (Forns)			
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Low risk	

Vallet-Pichard 2007

Study characteristics	
Patient Sampling	Study design: cross-sectional diagnostic test accuracy study Method of sampling: cohort-based Direction of data collection: not reported Country: France Inclusion criteria: anti HCV- and HCV-RNA–positive, liver biopsy prior to any antiviral therapy, laboratory assessments allowing FIB-4 calculation performed on the same day as liver biopsy or on the preceding day Exclusion criteria: HIV and HBV infection, alcohol excess, haemochromatosis, Wilson's disease, alpha1-antitrypsin deficiency, autoimmune hepatitis, non-alcoholic steatohepatitis, immunosuppression
Patient characteristics and setting	Centre details: single centre Sample size: 847 Mean age: 44 Gender (% male): 54.4 Mean BMI: not reported Mean ALT: 89
Index tests	Test name(s): FIB-4

Vallet-Pichard 2007 *(Continued)*

	Threshold(s) used: 1.45, 3.25
Target condition and reference standard(s)	Target condition(s): F3+ Reference standard: liver biopsy Quality of liver biopsy: not reported
Flow and timing	Flow: all participants received the index test and none were excluded from the analysis Time between index test and biopsy: same day
Comparative	Comparators: FibroTest
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Low risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index test (FIB-4)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index test (Forns)			
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	

Vallet-Pichard 2007 *(Continued)*

Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Low risk	

Wang 2015

Study characteristics			
Patient Sampling	Study design: cross-sectional diagnostic test accuracy study Method of sampling: cohort-based Direction of data collection: retrospective Country: Taiwan Inclusion criteria: adults with CHC who received percutaneous liver biopsy prior to antiviral treatment Exclusion criteria: alcoholism or use of hepatotoxic drug, HBV, HIV, HCC, liver transplant, recent platelet transfusion		
Patient characteristics and setting	Centre details: multicentre (2 centre) Sample size: 1473 Mean age: 54.4 Gender (% male): 53.2 Mean BMI: not reported Mean ALT: 134 (median)		
Index tests	Test name(s): FIB-4 Threshold(s) used: 0.75 2.15 1.45 3.25 2.00 6.50		
Target condition and reference standard(s)	Target condition(s): F2+, F3+, F4 Reference standard: liver biopsy Quality of liver biopsy: not reported		
Flow and timing	Flow: all participants received the index test and none were excluded from the analysis Time between index test and biopsy: same day		
Comparative	Comparators: APRI		
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient selection			

Wang 2015 *(Continued)*

Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Low risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index test (FIB-4)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index test (Forns)			
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Low risk	

Wang 2017

Study characteristics			

Wang 2017 *(Continued)*

Patient Sampling	Study design: cross-sectional diagnostic test accuracy study Method of sampling: cohort-based Direction of data collection: retrospective Country: Taiwan Inclusion criteria: adults with CHC or CHB Exclusion criteria: not reported
Patient characteristics and setting	Centre details: two centres Sample size: 1284 Mean age: 56 Gender (% male): 48.4 Mean BMI: not reported Median ALT: 77
Index tests	Test name(s): FIB-4 Threshold(s) used: 3.8
Target condition and reference standard(s)	Target condition(s): F4 Reference standard: liver biopsy Quality of liver biopsy: >1.5 cm but number of portal tracts not reported
Flow and timing	Flow: all participants received the index test and none were excluded from the analysis Time between index test and biopsy: 7 days
Comparative	Comparators: platelets, APRI, AAR; AAR/platelet ratio index (AARPRI), Lok, FibroQ, AP index, Pohl score
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Low risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index test (FIB-4)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		

Wang 2017 *(Continued)*

Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index test (Forns)			
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Low risk	

Yen 2018

Study characteristics	
Patient Sampling	Study design: cross-sectional diagnostic test accuracy study Method of sampling: cohort-based Direction of data collection: retrospective Country: Taiwan Inclusion criteria: HCV+ patients, treatment-naïve Exclusion criteria: the presence of other causes of liver disease, HCC, prior interferon therapy, human immunodeficiency virus (HIV) co-infection, and liver transplantation prior to liver biopsy
Patient characteristics and setting	Centre details: single centre, Kaoshiung Chang Gung Memorial Hospital, Taiwan Sample size: 1716 Mean age: 52.5 Gender (% male): 53.2 Mean BMI: 24.6 Mean ALT: 134
Index tests	Test name(s): FIB-4

Yen 2018 *(Continued)*

	Threshold(s) used: 2.9, 3.1
Target condition and reference standard(s)	Target condition(s): F2+, F3+, F4 Reference standard: liver biopsy Quality of liver biopsy: not reported
Flow and timing	Flow: all participants received the index test and none were excluded from the analysis Time between index test and biopsy: at time of biopsy
Comparative	Comparators: APRI
Notes	

Methodological quality

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Low risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index test (FIB-4)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index test (Forns)			
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	

Yen 2018 *(Continued)*

Are there concerns that the target condition as defined by the reference standard does not match the question?		Low concern
DOMAIN 4: Flow and timing		
Was there an appropriate interval between index test and reference standard?	Yes	
Did all patients receive the same reference standard?	Yes	
Were all patients included in the analysis?	Yes	
Could the patient flow have introduced bias?		Low risk

Yilmaz 2021

Study characteristics	
Patient Sampling	Study design: cross-sectional diagnostic test accuracy study Method of sampling: cohort-based Direction of data collection: retrospective Country: Turkey Inclusion criteria: chronic hepatitis C infection and liver biopsy result. Exclusion criteria: < 18 years old, previous HCV treatment, HBV, HDV, biopsy sample < 15 mm, missing laboratory values, autoimmune hepatitis, liver cancer, primary or secondary biliary cholangitis, primary sclerosing cholangitis, Wilson's disease, diabetes and active infection
Patient characteristics and setting	Centre details: multicentre (3 tertiary centres) Sample size: 114 Mean age: 63.6 Gender (% male): 19.3 Mean BMI: not reported Mean ALT: not reported
Index tests	Test name(s): FIB-4 Threshold(s) used: 1.45, 3.25
Target condition and reference standard(s)	Target condition(s): F2+, F3+, F4 Reference standard: liver biopsy Quality of liver biopsy: length > 15 mm but number of portal tracts not reported
Flow and timing	Flow: all participants received the index test and none were excluded from the analysis Time between index test and biopsy: not reported
Comparative	Comparators: APRI, ABA (age, bilirubin and albumin) index
Notes	
Methodological quality	

Yilmaz 2021 *(Continued)*

Item	Authors' judgement	Risk of bias	Applicability concerns
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	Unclear		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Unclear risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index test (FIB-4)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	Yes		
Could the conduct or interpretation of the index test have introduced bias?		Low risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index test (Forns)			
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	Unclear		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Unclear		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Unclear risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and timing			
Was there an appropriate interval between index test and reference standard?	Unclear		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Unclear risk	

Şirli 2010

Study characteristics			
Patient Sampling	Study design: cross-sectional diagnostic test accuracy study Method of sampling: cohort-based Direction of data collection: retrospective Country: Romania Inclusion criteria: chronic hepatitis C Exclusion criteria: HBV, chronic alcohol abuse, cholestatic chronic hepatitis, non-alcoholic steatohepatitis, autoimmune chronic hepatitis, haemochromatosis, Wilson's disease		
Patient characteristics and setting	Centre details: single centre Sample size: 150 Mean age: 50.4 Gender (% male): 32% Mean BMI: not reported Mean ALT: not reported		
Index tests	Test name(s): FIB-4 and Forns index Threshold(s) used: FIB-4 2.1365 (F2+) 2.3122 (F4). Forns index 4.57 (F2+) 5.93 (F4)		
Target condition and reference standard(s)	Target condition(s): F2+ and F4 Reference standard: liver biopsy Quality of liver biopsy: fragments of at least 2 cm, including at least 8 portal tracts, were considered adequate		
Flow and timing	Flow: all participants received the index test and none were excluded from the analysis Time between index test and biopsy: same day		
Comparative	Comparators: APRI, platelets, Lok, LSM		
Notes			
Methodological quality			
Item	**Authors' judgement**	**Risk of bias**	**Applicability concerns**
DOMAIN 1: Patient selection			
Was a consecutive or random sample of patients enrolled?	Yes		
Was a case-control design avoided?	Yes		
Did the study avoid inappropriate exclusions?	Yes		
Could the selection of patients have introduced bias?		Low risk	
Are there concerns that the included patients and setting do not match the review question?			Low concern
DOMAIN 2: Index test (FIB-4)			

Şirli 2010 *(Continued)*

Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 2: Index test (Forns)			
Were the index test results interpreted without knowledge of the results of the reference standard?	Yes		
If a threshold was used, was it pre-specified?	No		
Could the conduct or interpretation of the index test have introduced bias?		High risk	
Are there concerns that the index test, its conduct, or interpretation differ from the review question?			Low concern
DOMAIN 3: Reference standard			
Is the reference standard likely to correctly classify the target condition?	Yes		
Were the reference standard results interpreted without knowledge of the results of the index tests?	Yes		
Could the reference standard, its conduct, or its interpretation have introduced bias?		Low risk	
Are there concerns that the target condition as defined by the reference standard does not match the question?			Low concern
DOMAIN 4: Flow and timing			
Was there an appropriate interval between index test and reference standard?	Yes		
Did all patients receive the same reference standard?	Yes		
Were all patients included in the analysis?	Yes		
Could the patient flow have introduced bias?		Low risk	

AAR: aspartate to alanine aminotransferase ratio; **AIH:** autoimmune hepatitis; **ALT:** alanine transaminase; **AP index:** age-platelet index; **APASL:** Asian Pacific Association for the Study of the Liver; **APRI:** AST to Platelet Ratio Index; **AST:** aspartate aminotransferase; **BMI:** body mass index; **CDS:** (the Bonacini) cirrhosis discriminant score; **CHC:** chronic hepatitis C; **ElastPQ:** elastography point quantification; **ELF:** Enhanced Liver Fibrosis; **ELISA:** enzyme-linked immunosorbent assay; **FIB-4:** Fibrosis-4; **GUCI:** Göteborg University Cirrhosis Index; **HAV:** hepatitis A virus; **HBV:** hepatitis B virus; **HCC:** hepatocellular carcinoma; **HCV:** hepatitis C virus; **IFN:** interferon; **LSM:** liver stiffness measurement; **NAFLD:** non-alcoholic fatty liver disease; **PCR:** polymerase chain reaction; **RNA:** ribonucleic acid; **TE:** transient elastography; **ULN:** upper limit of normal

D A T A

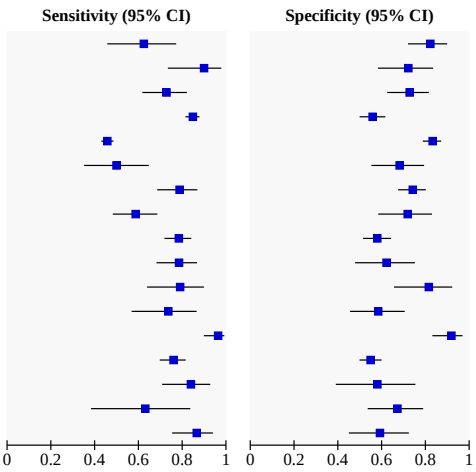
Presented below are all the data for all of the tests entered into the review.

Table Tests. Data tables by test

Test	No. of studies	No. of participants
1 FIB4 1.45 - F2	17	5098
2 FIB4 3.25 - F2	5	2673
3 FIB4 1.45 - F3	39	86907
4 FIB4 3.25 - F3	24	81430
5 FIB4 1.45 - F4	10	4537
6 FIB4 3.25 - F4	9	5075
7 Forns 4.2 - F2	17	4402
8 Forns 6.9 - F2	12	3287
9 Forns 4.2 - F3	1	180
10 Forns 6.9 - F3	3	1516
11 Forns 4.2 - F4	1	150
12 Forns 6.9 - F4	2	234

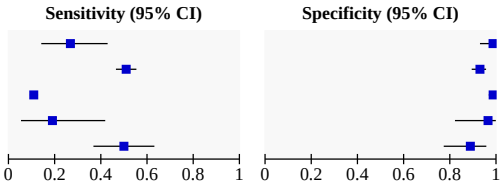
Test 1. FIB4 1.45 - F2

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Amorim 2012	25	14	15	65	0.63 [0.46, 0.77]	0.82 [0.72, 0.90]
Andrés-Otero 2016	27	15	3	39	0.90 [0.73, 0.98]	0.72 [0.58, 0.84]
Conti 2019	59	25	22	67	0.73 [0.62, 0.82]	0.73 [0.63, 0.82]
De Oliveira 2016	433	127	77	161	0.85 [0.81, 0.88]	0.56 [0.50, 0.62]
Ferenci 2014	581	54	687	270	0.46 [0.43, 0.49]	0.83 [0.79, 0.87]
Fontaine 2009	24	20	24	43	0.50 [0.35, 0.65]	0.68 [0.55, 0.79]
Gounder 2018	67	51	18	147	0.79 [0.69, 0.87]	0.74 [0.68, 0.80]
Guilabert 2010	57	16	40	41	0.59 [0.48, 0.69]	0.72 [0.58, 0.83]
Ladero 2010	146	102	40	141	0.78 [0.72, 0.84]	0.58 [0.52, 0.64]
Ozel 2015	66	20	18	33	0.79 [0.68, 0.87]	0.62 [0.48, 0.75]
Paranaguá-Vezozzo 2017	34	7	9	31	0.79 [0.64, 0.90]	0.82 [0.66, 0.92]
Schmid 2015	28	27	10	38	0.74 [0.57, 0.87]	0.58 [0.46, 0.71]
Shaikh 2009	81	6	3	68	0.96 [0.90, 0.99]	0.92 [0.83, 0.97]
Shiha 2017	162	176	51	215	0.76 [0.70, 0.82]	0.55 [0.50, 0.60]
Trang 2008	42	13	8	18	0.84 [0.71, 0.93]	0.58 [0.39, 0.75]
Usluer 2012	12	19	7	39	0.63 [0.38, 0.84]	0.67 [0.54, 0.79]
Yilmaz 2021	52	22	8	32	0.87 [0.75, 0.94]	0.59 [0.45, 0.72]



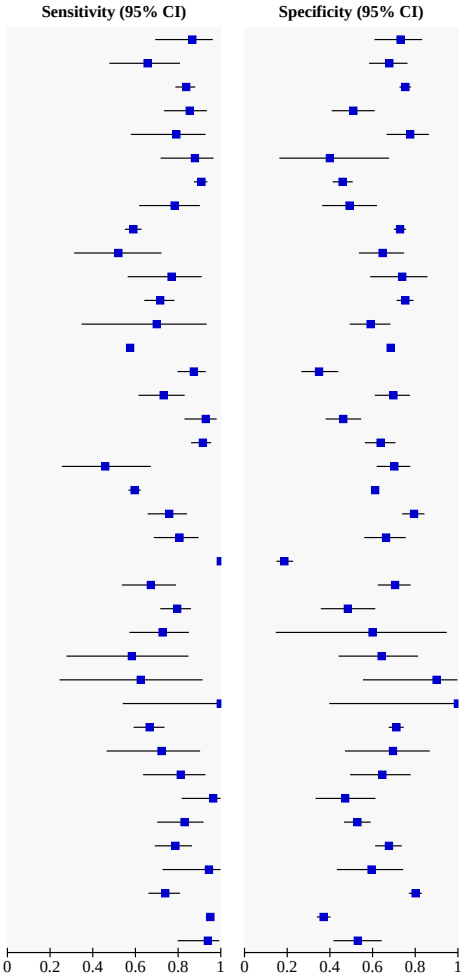
Test 2. FIB4 3.25 - F2

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Amorim 2012	11	1	30	77	0.27 [0.14, 0.43]	0.99 [0.93, 1.00]
De Oliveira 2016	260	20	250	268	0.51 [0.47, 0.55]	0.93 [0.89, 0.96]
Ferenci 2014	139	4	1129	320	0.11 [0.09, 0.13]	0.99 [0.97, 1.00]
Patel 2017	4	1	17	28	0.19 [0.05, 0.42]	0.97 [0.82, 1.00]
Yilmaz 2021	30	6	30	48	0.50 [0.37, 0.63]	0.89 [0.77, 0.96]



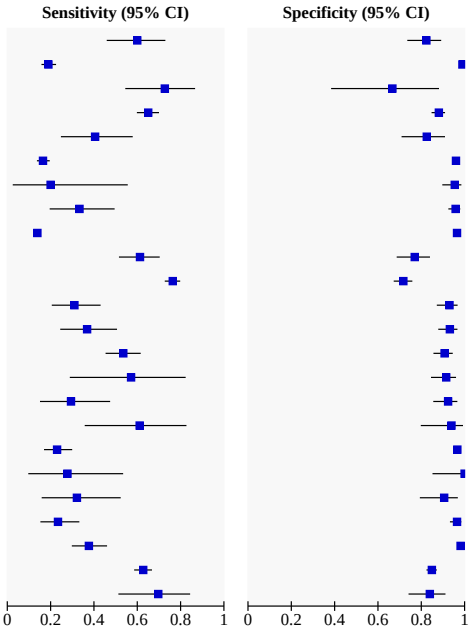
Test 3. FIB4 1.45 - F3

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Abdel-Hameed 2021a	26	18	4	49	0.87 [0.69, 0.96]	0.73 [0.61, 0.83]
Abdel-Hameed 2021b	23	36	12	76	0.66 [0.48, 0.81]	0.68 [0.58, 0.76]
Abdelsameea 2020	217	249	42	762	0.84 [0.79, 0.88]	0.75 [0.73, 0.78]
Ahmad 2011	47	50	8	52	0.85 [0.73, 0.94]	0.51 [0.41, 0.61]
Alboraie 2015	19	17	5	59	0.79 [0.58, 0.93]	0.78 [0.67, 0.86]
Baldwin 2020	29	9	4	6	0.88 [0.72, 0.97]	0.40 [0.16, 0.68]
De Oliveira 2016	320	241	32	205	0.91 [0.87, 0.94]	0.46 [0.41, 0.51]
Demma 2018	29	32	8	31	0.78 [0.62, 0.90]	0.49 [0.36, 0.62]
Ferenci 2014	376	259	261	696	0.59 [0.55, 0.63]	0.73 [0.70, 0.76]
Fontaine 2009	13	30	12	55	0.52 [0.31, 0.72]	0.65 [0.54, 0.75]
Fouad 2018	20	12	6	34	0.77 [0.56, 0.91]	0.74 [0.59, 0.86]
Gamil 2017	121	119	48	364	0.72 [0.64, 0.78]	0.75 [0.71, 0.79]
Gökan 2016	7	45	3	65	0.70 [0.35, 0.93]	0.59 [0.49, 0.68]
Hashem 2021	6679	18110	4936	39382	0.58 [0.57, 0.58]	0.68 [0.68, 0.69]
Hseih 2012	97	82	14	44	0.87 [0.80, 0.93]	0.35 [0.27, 0.44]
Loko 2008	52	39	19	90	0.73 [0.61, 0.83]	0.70 [0.61, 0.78]
Maheshwari 2013	53	79	4	68	0.93 [0.83, 0.98]	0.46 [0.38, 0.55]
Martinez 2011	142	67	13	118	0.92 [0.86, 0.95]	0.64 [0.56, 0.71]
Portilla 2009	11	42	13	99	0.46 [0.26, 0.67]	0.70 [0.62, 0.78]
Shiha 2022a	690	1653	467	2607	0.60 [0.57, 0.62]	0.61 [0.60, 0.63]
Shiha 2022b	69	51	22	197	0.76 [0.66, 0.84]	0.79 [0.74, 0.84]
Shiha 2022c	50	33	12	65	0.81 [0.69, 0.90]	0.66 [0.56, 0.76]
Shiha 2022d	390	331	0	76	1.00 [0.99, 1.00]	0.19 [0.15, 0.23]
Shiha 2022e	39	43	19	103	0.67 [0.54, 0.79]	0.71 [0.62, 0.78]
Shiha 2022f	105	33	27	31	0.80 [0.72, 0.86]	0.48 [0.36, 0.61]
Shiha 2022g	32	2	12	3	0.73 [0.57, 0.85]	0.60 [0.15, 0.95]
Shiha 2022h	7	10	5	18	0.58 [0.28, 0.85]	0.64 [0.44, 0.81]
Shiha 2022i	5	1	3	9	0.63 [0.24, 0.91]	0.90 [0.55, 1.00]
Shiha 2022j	6	0	0	4	1.00 [0.54, 1.00]	1.00 [0.40, 1.00]
Sterling 2006	116	189	58	467	0.67 [0.59, 0.74]	0.71 [0.68, 0.75]
Stibbe 2011	13	7	5	16	0.72 [0.47, 0.90]	0.70 [0.47, 0.87]
Tanwar 2017	26	17	6	31	0.81 [0.64, 0.93]	0.65 [0.49, 0.78]
Trang 2008	27	28	1	25	0.96 [0.82, 1.00]	0.47 [0.33, 0.61]
Trifan 2009	44	122	9	137	0.83 [0.70, 0.92]	0.53 [0.47, 0.59]
Tural 2009	74	74	20	155	0.79 [0.69, 0.86]	0.68 [0.61, 0.74]
Udompap 2020	17	17	1	25	0.94 [0.73, 1.00]	0.60 [0.43, 0.74]
Vallet-Pichard 2007	108	139	38	562	0.74 [0.66, 0.81]	0.80 [0.77, 0.83]
Wang 2015	526	579	27	341	0.95 [0.93, 0.97]	0.37 [0.34, 0.40]
Yilmaz 2021	31	38	2	43	0.94 [0.80, 0.99]	0.53 [0.42, 0.64]



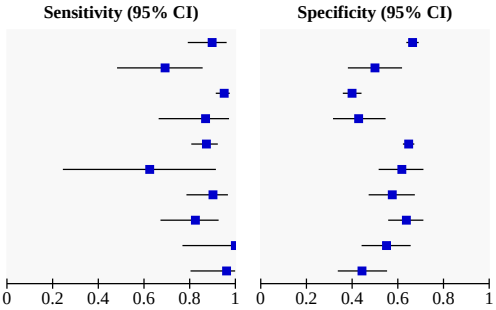
Test 4. FIB4 3.25 - F3

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Ahmad 2011	33	18	22	84	0.60 [0.46, 0.73]	0.82 [0.74, 0.89]
Attallah 2012	104	27	444	2637	0.19 [0.16, 0.23]	0.99 [0.99, 0.99]
Baldwin 2020	24	5	9	10	0.73 [0.54, 0.87]	0.67 [0.38, 0.88]
De Oliveira 2016	229	53	123	393	0.65 [0.60, 0.70]	0.88 [0.85, 0.91]
Demma 2018	15	11	22	52	0.41 [0.25, 0.58]	0.83 [0.71, 0.91]
Ferenci 2014	105	38	532	917	0.16 [0.14, 0.20]	0.96 [0.95, 0.97]
Gökan 2016	2	5	8	105	0.20 [0.03, 0.56]	0.95 [0.90, 0.99]
Gounder 2018	14	10	28	231	0.33 [0.20, 0.50]	0.96 [0.93, 0.98]
Hashem 2021	1614	2012	10001	55480	0.14 [0.13, 0.15]	0.97 [0.96, 0.97]
Hseih 2012	68	29	43	97	0.61 [0.52, 0.70]	0.77 [0.69, 0.84]
Ikatura 2021	437	125	135	317	0.76 [0.73, 0.80]	0.72 [0.67, 0.76]
Loko 2008	22	9	49	120	0.31 [0.21, 0.43]	0.93 [0.87, 0.97]
Maheshwari 2013	21	10	36	137	0.37 [0.24, 0.51]	0.93 [0.88, 0.97]
Martinez 2011	83	17	72	167	0.54 [0.45, 0.62]	0.91 [0.86, 0.95]
Qian 2019	8	9	6	97	0.57 [0.29, 0.82]	0.92 [0.84, 0.96]
Schmoyer 2020	10	8	24	97	0.29 [0.15, 0.47]	0.92 [0.86, 0.97]
Silva Junior 2014	11	2	7	31	0.61 [0.36, 0.83]	0.94 [0.80, 0.99]
Sterling 2006	40	22	134	634	0.23 [0.17, 0.30]	0.97 [0.95, 0.98]
Stibbe 2011	5	0	13	23	0.28 [0.10, 0.53]	1.00 [0.85, 1.00]
Trang 2008	9	5	19	48	0.32 [0.16, 0.52]	0.91 [0.79, 0.97]
Tural 2009	22	8	72	221	0.23 [0.15, 0.33]	0.97 [0.93, 0.98]
Vallet-Pichard 2007	55	13	91	688	0.38 [0.30, 0.46]	0.98 [0.97, 0.99]
Wang 2015	347	139	206	781	0.63 [0.59, 0.67]	0.85 [0.82, 0.87]
Yilmaz 2021	23	13	10	68	0.70 [0.51, 0.84]	0.84 [0.74, 0.91]



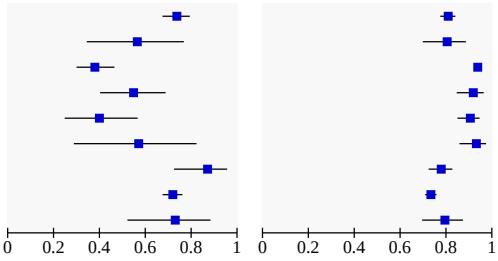
Test 5. FIB4 1.45 - F4

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Abdelsameea 2020	53	406	6	805	0.90 [0.79, 0.96]	0.66 [0.64, 0.69]
Abo El-Khair 2019	18	37	8	37	0.69 [0.48, 0.86]	0.50 [0.38, 0.62]
De Oliveira 2016	214	344	11	229	0.95 [0.91, 0.98]	0.40 [0.36, 0.44]
Demma 2018	20	44	3	33	0.87 [0.66, 0.97]	0.43 [0.32, 0.55]
Ferenci 2014	124	511	18	939	0.87 [0.81, 0.92]	0.65 [0.62, 0.67]
Fontaine 2009	5	39	3	63	0.63 [0.24, 0.91]	0.62 [0.52, 0.71]
Güzelbulut 2011	46	42	5	57	0.90 [0.79, 0.97]	0.58 [0.47, 0.67]
Loko 2008	33	58	7	102	0.82 [0.67, 0.93]	0.64 [0.56, 0.71]
Schmid 2015	14	40	0	49	1.00 [0.77, 1.00]	0.55 [0.44, 0.66]
Yilmaz 2021	25	49	1	39	0.96 [0.80, 1.00]	0.44 [0.34, 0.55]



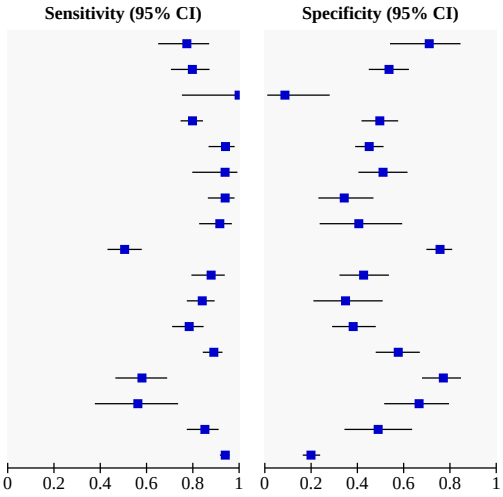
Test 6. FIB4 3.25 - F4

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
De Oliveira 2016	166	109	59	464	0.74 [0.68, 0.79]	0.81 [0.78, 0.84]
Demma 2018	13	15	10	62	0.57 [0.34, 0.77]	0.81 [0.70, 0.89]
Ferenci 2014	54	89	88	1361	0.38 [0.30, 0.47]	0.94 [0.93, 0.95]
Güzelbulut 2011	28	8	23	91	0.55 [0.40, 0.69]	0.92 [0.85, 0.96]
Loko 2008	16	15	24	145	0.40 [0.25, 0.57]	0.91 [0.85, 0.95]
Schmid 2015	8	6	6	83	0.57 [0.29, 0.82]	0.93 [0.86, 0.97]
Tsukano 2017	34	58	5	205	0.87 [0.73, 0.96]	0.78 [0.72, 0.83]
Yen 2018	314	340	122	940	0.72 [0.68, 0.76]	0.73 [0.71, 0.76]
Yilmaz 2021	19	18	7	70	0.73 [0.52, 0.88]	0.80 [0.70, 0.87]



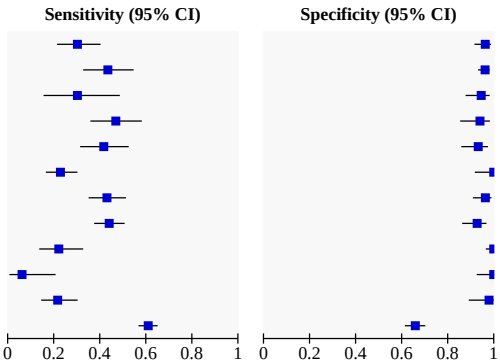
Test 7. Forns 4.2 - F2

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Alboraie 2015	48	11	14	27	0.77 [0.65, 0.87]	0.71 [0.54, 0.85]
Bourliere 2006	79	63	20	73	0.80 [0.71, 0.87]	0.54 [0.45, 0.62]
Corradi 2009	13	21	0	2	1.00 [0.75, 1.00]	0.09 [0.01, 0.28]
Crisan 2012	226	82	57	81	0.80 [0.75, 0.84]	0.50 [0.42, 0.58]
Forns 2002a	80	146	5	120	0.94 [0.87, 0.98]	0.45 [0.39, 0.51]
Forns 2002b	31	45	2	47	0.94 [0.80, 0.99]	0.51 [0.40, 0.62]
Güzelbulut 2011	78	44	5	23	0.94 [0.86, 0.98]	0.34 [0.23, 0.47]
Koller 2014	66	19	6	13	0.92 [0.83, 0.97]	0.41 [0.24, 0.59]
Ladero 2010	94	59	92	184	0.51 [0.43, 0.58]	0.76 [0.70, 0.81]
Leroy 2007	80	51	11	38	0.88 [0.79, 0.94]	0.43 [0.32, 0.54]
Loko 2008	132	28	25	15	0.84 [0.77, 0.89]	0.35 [0.21, 0.51]
Macías 2006	120	68	33	42	0.78 [0.71, 0.85]	0.38 [0.29, 0.48]
Martinez 2011	204	47	25	64	0.89 [0.84, 0.93]	0.58 [0.48, 0.67]
Mossong 2011	47	24	34	81	0.58 [0.47, 0.69]	0.77 [0.68, 0.85]
Sebastiani 2008a	18	16	14	32	0.56 [0.38, 0.74]	0.67 [0.52, 0.80]
Sebastiani 2008b	98	25	17	24	0.85 [0.77, 0.91]	0.49 [0.34, 0.64]
Sebastiani 2012	519	369	33	92	0.94 [0.92, 0.96]	0.20 [0.16, 0.24]



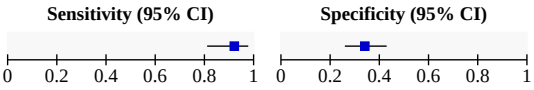
Test 8. Forns 6.9 - F2

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Bourliere 2006	30	5	69	131	0.30 [0.21, 0.40]	0.96 [0.92, 0.99]
Forns 2002a	37	10	48	256	0.44 [0.33, 0.55]	0.96 [0.93, 0.98]
Forns 2002b	10	5	23	87	0.30 [0.16, 0.49]	0.95 [0.88, 0.98]
Güzelbulut 2011	39	4	44	63	0.47 [0.36, 0.58]	0.94 [0.85, 0.98]
Leroy 2007	38	6	53	83	0.42 [0.32, 0.53]	0.93 [0.86, 0.97]
Loko 2008	36	0	121	43	0.23 [0.17, 0.30]	1.00 [0.92, 1.00]
Macías 2006	66	4	87	106	0.43 [0.35, 0.51]	0.96 [0.91, 0.99]
Martinez 2011	101	8	128	103	0.44 [0.38, 0.51]	0.93 [0.86, 0.97]
Mossong 2011	18	0	63	105	0.22 [0.14, 0.33]	1.00 [0.97, 1.00]
Sebastiani 2008a	2	0	30	48	0.06 [0.01, 0.21]	1.00 [0.93, 1.00]
Sebastiani 2008b	25	1	90	48	0.22 [0.15, 0.30]	0.98 [0.89, 1.00]
Sebastiani 2012	337	157	215	304	0.61 [0.57, 0.65]	0.66 [0.61, 0.70]



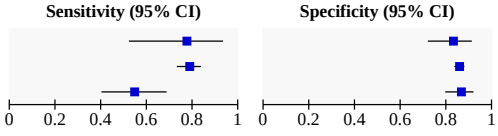
Test 9. Forns 4.2 - F3

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Leroy 2007	47	85	4	44	0.92 [0.81, 0.98]	0.34 [0.26, 0.43]



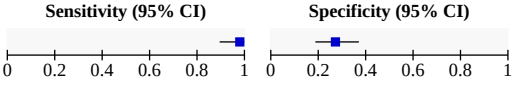
Test 10. Forns 6.9 - F3

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Andrés-Otero 2016	14	11	4	55	0.78 [0.52, 0.94]	0.83 [0.72, 0.91]
Iacobellis 2005	192	141	51	868	0.79 [0.73, 0.84]	0.86 [0.84, 0.88]
Leroy 2007	28	17	23	112	0.55 [0.40, 0.69]	0.87 [0.80, 0.92]



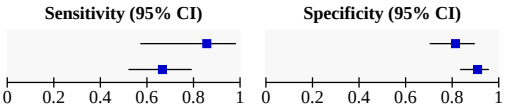
Test 11. Forns 4.2 - F4

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Güzelbulut 2011	50	72	1	27	0.98 [0.90, 1.00]	0.27 [0.19, 0.37]



Test 12. Forns 6.9 - F4

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Andrés-Otero 2016	12	13	2	57	0.86 [0.57, 0.98]	0.81 [0.70, 0.90]
Güzelbulut 2011	34	9	17	90	0.67 [0.52, 0.79]	0.91 [0.83, 0.96]



ADDITIONAL TABLES

Table 1. Translation of different fibrosis staging systems to METAVIR

Ishak (Ishak 1995)	Knodell (Knodell 1981)	Scheuer (Scheuer 1991)	METAVIR (METAVIR 1996)
0	0	0	0
1	1	1	1
2, 3	1	2	2
4, 5	3	3	3
6	4	4	4

Table 2. Pooled estimates

Test - target condition	Cut-off	Number of studies	Median prevalence	Sensitivity (95% CI)	Specificity (95% CI)	LR+ (95% CI)	LR- (95% CI)
FIB-4 - F2	1.45	17	46.8%	76.2% (from 68.9% to 82.3%)	70.0% (from 64.0% to 75.4%)	2.5 (from 2.1 to 3.1)	0.34 (from 0.25 to 0.45)
FIB-4 - F2	3.25	5	52.6%	29.2% (from 15.8% to 47.6%)	96.6% (from 92.6% to 98.5%)	8.7 (from 5.5 to 13.7)	0.73 (from 0.60 to 0.90)
FIB-4 - F3	1.45	39	30.9%	81.1% (from 75.6% to 85.6%)	62.3% (from 57.4% to 66.9%)	2.2 (from 2 to 2.4)	0.30 (from 0.24 to 0.38)
FIB-4 - F3	3.25	24	34.8%	41.4% (from 33.0% to 50.4%)	92.6% (from 89.5% to 94.9%)	5.6 (from 4.4 to 7.1)	0.63 (from 0.56 to 0.72)
FIB-4 - F4	1.45	10	21.4%	89.1% (from 83.9% to 92.8%)	55.6% (from 49.3% to 61.7%)	2.0 (from 1.8 to 2.3)	0.20 (from 0.14 to 0.28)
FIB-4 - F4	3.25	9	22.8%	61.2% (from 50.7% to 70.8%)	85.9% (from 80.2% to 90.2%)	4.4 (from 3.4 to 5.5)	0.45 (from 0.36 to 0.56)
Forns index - F2	4.2	17	54.5%	84.7% (from 77.9% to 89.7%)	47.9% (from 38.6% to 57.3%)	1.6 (from 1.4 to 1.9)	0.32 (from 0.25 to 0.41)
Forns index - F2	6.9	12	52.5%	34.1% (from 26.4% to 42.8%)	97.3% (from 92.9% to 99.0%)	12.5 (from 5.7 to 27.2)	0.68 (from 0.61 to 0.75)
Forns index - F3	6.9	3	21.4%	-	-	-	-
Forns index - F4	6.9	2	25.3%	-	-	-	-

The bivariate model has been fitted only when at least four studies were included.
CI: confidence interval; **LR+:** positive likelihood ratio; **LR-:** negative likelihood ratio

Table 3. Comparisons

F2 - Indirect comparisons	Number of studies	Relative sensitivity (95% CI)	Relative specificity (95% CI)	P value[a]
Forns index cut-off 4.2 vs FIB-4 cut-off 1.45	34	1.12 (from 1.00 to 1.25)	0.69 (from 0.57 to 0.84)	0.004
Forns index cut-off 6.9 vs FIB-4 cut-off 3.25	17	1.14 (from 0.66 to 1.96)	0.99 (from 0.95 to 1.04)	0.975

CI: confidence interval; **vs:** versus
[a]P value refers to the difference in overall accuracy between the two tests.

Table 4. Sensitivity analyses

Test – target condition	Cut-off	Low risk of bias	Number of studies at low risk of bias / Number of included studies	Sensitivity (95% CI)	Specificity (95% CI)
FIB-4 – F2	1.45	Participant selection	14/17	76.0% (from 67.9% to 82.6%)	71.0% (from 65.2% to 77.8%)
		Index test	12/17	76.2% (from 66.4% to 83.4%)	69.2% (from 61.0% to 76.3%)
		Reference standard	2/17	-	-
		Participant flow	7/17	78.6% (from 74.4% to 82.2%)	67.3% (from 60.8% to 73.2%)
FIB-4 – F2	3.25	Participant selection	3/5	-	-
		Index test	5/5	29.2% (from 15.8% to 47.6%)	96.6% (from 92.6% to 98.5%)
		Reference standard	0/5	-	-
		Participant flow	0/5	-	-
FIB-4 – F3	1.45	Participant selection	16/39	82.1% (from 74.9% to 87.5%)	61.0% (from 53.7% to 67.8%)
		Index test	29/39	80.8% (from 74.0% to 86.2%)	62.6% (from 56.6% to 68.2%)

Table 4. Sensitivity analyses *(Continued)*

		Reference standard	13/39	80.6% (from 65.0% to 90.3%)	64.6% (from 54.2% to 73.7%)
		Participant flow	11/39	82.5% (from 75.8% to 87.6%)	64.5% (from 54.9% to 73.0%)
FIB-4 – F3	3.25	Participant selection	14/24	39.4% (from 30.1% to 49.5%)	93.2% (from 89.7% to 95.5%)
		Index test	20/24	39.3% (from 30.6% to 48.7%)	93.2% (from 90.0% to 95.5%)
		Reference standard	1/24	-	-
		Participant flow	9/24	37.7% (from 27.8% to 48.8%)	95.4% (from 91.1% to 97.7%)
FIB-4 – F4	1.45	Participant selection	4/10	90.1% (from 53.0% to 98.6%)	54.2% (from 16.2% to 87.9%)
		Index test	9/10	90.6% (from 86.8% to 93.4%)	56.0% (from 49.2% to 62.6%)
		Reference standard	1/10	-	-
		Participant flow	5/10	85.8% (from 78.0% to 91.2%)	60.4% (from 54.4% to 66.2%)
FIB-4 – F4	3.25	Participant selection	6/9	55.2% (from 43.0% to 66.8%)	87.9% (from 80.6% to 92.7%)
		Index test	7/9	57.2% (from 45.6% to 68.0%)	88.5% (from 83.9% to 92.0%)
		Reference standard	1/9	-	-
		Participant flow	5/9	61.4% (from 46.1% to 74.8%)	86.4% (from 77.7% to 92.1%)
Forns index – F2	4.2	Participant selection	15/17	83.6% (from 76.8% to 88.8%)	49.2% (from 40.6% to 57.8%)
		Index test	15/17	85.6% (from 78.0% to 90.9%)	46.4% (from 36.3% to 56.7%)
		Reference standard	4/17	92.7% (from 74.2% to 98.2%)	47.7% (from 32.7% to 63.1%)

Table 4. Sensitivity analyses *(Continued)*

		Participant flow	10/17	84.3% (from 75.2% to 90.5%)	48.7% (from 37.3% to 60.2%)
Forns index – F2	6.9	Participant selection	12/12	34.1% (from 26.4% to 42.8%)	97.3% (from 92.9% to 99.0%)
		Index test	12/12	34.1% (from 26.4% to 42.8%)	97.3% (from 92.9% to 99.0%)
		Reference standard	2/12	-	-
		Participant flow	8/12	33.3% (from 22.7% to 45.9%)	97.2% (from 89.7% to 99.3%)
FIB-4 – F2	1.45	Only full text	16/17	77.5% (from 70.5% to 83.2%)	70.1% (from 63.7% to 75.8%)
FIB-4 – F2	3.25	Only full text	4/5	33.3% (from 22.7% to 45.9%)	97.2% (from 89.7% to 99.3%)
FIB-4 – F3	1.45	Only full text	34/39	80.9% (from 74.7% to 85.9%)	63.7% (from 58.5% to 68.7%)
FIB-4 – F3	3.25	Only full text	21/24	40.0% (from 31.1% to 49.7%)	93.3% (from 90.3% to 95.5%)
FIB-4 – F4	1.45	Only full text	8/10	90.1% (from 86.1% to 93.1%)	56.4% (from 49.4% to 63.1%)
FIB-4 – F4	3.25	Only full text	8/9	61.2% (from 49.4% to 71.8%)	86.3% (from 80.0% to 90.9%)
Forns index – F2	4.2	Only full text	17/17	85.3% (from 79.7% to 89.6%)	46.4% (from 38.1% to 55.0%)
Forns index – F2	6.9	Only full text	12/12	38.4% (from 28.1% to 49.8%)	95.6% (from 87.6% to 98.5%)

The bivariate model has been fitted only when at least four studies were included.
CI: confidence interval

Table 5. Subgroup analyses

Test – target condition	Cut-off	Subgroup	Number of studies	Sensitivity (95% CI)	Specificity (95% CI)	P value

Table 5. Subgroup analyses *(Continued)*

FIB-4 – F2	1.45	HIV YES	4	82.4% (from 60.3% to 93.5%)	73.5% (from 54.9% to 86.3%)	0.708
		HIV NO	13	72.8% (from 65.3% to 79.2%)	68.7% (from 62.7% to 74.1%)	
FIB-4 – F2	3.25	HIV YES	0	-	-	-
		HIV NO	5	29.2% (from 15.8% to 47.6%)	96.6% (from 92.6% to 98.5%)	
FIB-4 – F3	1.45	HIV YES	6	75.2% (from 59.7% to 86.1%)	67.3% (from 61.4% to 72.8%)	0.766
		HIV NO	33	82.37% (from 76.4% to 86.9%)	61.6% (from 55.9% to 66.9%)	
FIB-4 – F3	3.25	HIV YES	5	24.7% (from 19.9% to 30.2%)	96.5% (from 93.1% to 98.3%)	0.139
		HIV NO	19	46.1% (from 36.2% to 56.3%)	90.9% (from 87.2% to 93.7%)	
FIB-4 – F4	1.45	HIV YES	2	92.4% (from 51.8% to 99.3%)	59.8% (from 51.5% to 67.6%)	0.916
		HIV NO	8	88.9% (from 82.6% to 93.2%)	54.5% (from 47.0% to 61.7%)	
FIB-4 – F4	3.25	HIV YES	2	44.5% (from 31.6% to 58.2%)	91.6% (from 87.3% to 94.5%)	0.427
		HIV NO	7	65.1% (from 54.0% to 74.7%)	83.9% (from 76.9% to 89.1%)	
Forns index – F2	4.2	HIV YES	3	83.0% (from 78.1% to 87.1%)	39.3% (from 33.3% to 45.8%)	0.264
		HIV NO	14	85.2% (from 76.6% to 91.1%)	49.6% (from 38.4% to 60.8%)	
Forns index – F2	6.9	HIV YES	3	35.5% (from 24.9% to 47.7%)	97.4% (from 88.1% to 99.5%)	0.821

Table 5. Subgroup analyses *(Continued)*

		HIV NO	9	33.2% (from 23.6% to 44.5%)	97.2% (from 91.0% to 99.2%)	
FIB-4 – F2	1.45	ALT > 80	3	Model did not converge		-
		ALT < 80	10			
		Missing ALT	4	-	-	-
FIB-4 – F2	3.25	ALT > 80	1	Model did not converge		-
		ALT < 80	3			
		Missing ALT	1	-	-	-
FIB-4 – F3	1.45	ALT > 80	12	83.9% (from 76.2% to 89.4%)	59.1% (from 50.6% to 67.2%)	0.155
		ALT < 80	12	77.2% (from 68.0% to 84.3%)	69.9% (from 65.6% to 73.8%)	
		Missing ALT	15	-	-	-
FIB-4 – F3	3.25	ALT > 80	11	41.0% (from 29.7% to 53.4%)	92.6% (from 88.5% to 95.4%)	0.962
		ALT < 80	9	39.9% (from 25.3% to 56.6%)	93.9% (from 88.1% to 96.9%)	
		Missing ALT	4	-	-	-
FIB-4 – F4	1.45	ALT > 80	2	90.7% (from 76.8% to 96.7%)	51.6% (from 35.1% to 67.8%)	0.659
		ALT < 80	4	90.6% (from 84.1% to 94.6%)	59.7% (from 51.4% to 67.6%)	
		Missing ALT	4	-	-	-
FIB-4 – F4	3.25	ALT > 80	3	63.8% (from 46.8% to 77.9%)	82.7% (from 71.9% to 89.9%)	0.381
		ALT < 80	4	65.4% (from 43.2% to 82.4%)	87.6% (from 79.2% to 92.9%)	

Table 5. Subgroup analyses *(Continued)*

		Missing ALT	2	-	-	-
Forns index – F2	4.2	ALT > 80	8	85.8% (from 76.8% to 91.7%)	46.3% (from 34.6% to 58.5%)	0.980
		ALT < 80	8	86.6% (from 75.1% to 93.2%)	44.7% (from 32.2% to 57.9%)	
		Missing ALT	1	-	-	-
Forns index – F2	6.9	ALT > 80	7	38.1% (from 29.0% to 48.2%)	96.1% (from 88.7% to 98.7%)	0.746
		ALT < 80	4	29.0% (from 14.2% to 50.3%)	96.8% (from 90.0% to 99.0%)	
		Missing ALT	1	-	-	-

ALT: alanine aminotransferase; **CI:** confidence interval

APPENDICES

Appendix 1. Search strategies

We carried out the following searches on 13 April 2022.

Database	Date range	Search Strategy
The Cochrane Hepato-Biliary Diagnostic Test Accuracy Studies Register	January 2003 to 13 April 2022	(FIB-4 or FIB4 or fibrosis-4 or Forn*1) AND fibro* AND (hepatitis C or hep-C or CHC or HCV)
MEDLINE Ovid	January 2003 to 13 April 2022	1. exp Hepatitis C/ 2. ((hepatitis or hep) adj3 C).tw,ot,kf. 3. (CHC or HCV).tw,ot,kf. 4. or/1-3 5. Liver Cirrhosis/ 6. (Fibrosis/ or (fibrosis or fibroses or fibrotic or “liver disease*”).tw,ot,kf.) and (hepat* or liver*).mp 7. or/5-6 8. 4 and 7 9. (FIB-4 or FIB4 or fibrosis-4).tw,ot,kf. 10. Forn*1.tw,ot,kf. 11. or/9-10 12. *Biological Markers/bl, du 13. ((serum or blood or biochemical or biologic*) adj3 (marker* or biomarker*)).tw,ot,kf 14. ((indirect or non-invasiv* or noninvasive*) adj6 (marker* or test* or assay* or measurement* or assess* or method*)).tw,ot,kf 15. or/12-14

(Continued)

		16. di.fs. 17. du.fs. 18. or/16-17 19. 4 and 11 20. 7 and 11 21. 4 and 15 and 18 22. 8 and 15 23. or/19-22 24. animals/ not humans/ 25. 23 not 24
Embase Ovid	January 2003 to 13 April 2022	1. exp hepatitis C/ 2. ((hepatitis or hep) adj3 C).ti,ab,kw. 3. (CHC or HCV).ti,ab,kw. 4. or/1-3 5. liver cirrhosis/ 6. (Fibrosis/ or (fibrosis or fibroses or fibrotic or "liver disease*").ti,ab,kw) and (hepat* or liver*).ti,ab,kw 7. or/5-6 8. 4 and 7 9. (FIB-4 or FIB4 or fibrosis-4).ti,ab,kw. 10. Forn*1.ti,ab,kw. 11. or/9-10 12. ((serum or blood or biochemical or biologic*) adj3 (marker* or biomarker*)).ti,ab,kw 13. ((indirect or non-invasiv* or noninvasiv*) adj6 (marker* or test* or assay* or measurement* or assess* or method*)).ti,ab,kw 14. or/12-13 15. di.fs. 16. du.fs. 17. or/15-16 18. 11 and (4 or 7) 19. 4 and 14 and 17 20. 8 and 14 21. 18 or 19 or 20
CINAHL (EBSCOhost)	January 2003 to 13 April 2022	S1 TX "FIB 4" S2 TX FIB4 S3 TX "forns index" OR "forn's index" S4 S1 OR S2 OR S3
Science Citation Index Expanded (Web of Science)	January 2003 to 13 April 2022	TS=(FIB4 OR "FIB 4" OR "forns index" OR "forn's index")
LILACS (BIREME)	January 2003 to 13 April 2022	fib4 OR "fib 4" OR "forn's index" OR "forns index"

Appendix 2. Signalling questions and corresponding answers for the assessment of methodological quality using the QUADAS-2 tool

Quality assessed	Signalling question	Choice	Comment
Domain 1			

(Continued)

Participant sampling	Was a consecutive or random sample of participants enroled?	Yes/No/Unclear	Yes, if there was a consecutive or random sample of participants enroled. No, if this was not the case. Unclear if there was no information available.
	Was a case-control design avoided?	Yes/No/Unclear	Yes, if a case-control design was avoided. No, if it was not avoided. Unclear, if it was not stated.
	Did the study avoid inappropriate exclusions?	Yes/No/Unclear	For example, exclusion of people with severe or low fibrosis, obesity, etc. Yes, if the study avoided inappropriate exclusions. No, if it did not avoid inappropriate exclusions. Unclear, if no such information was provided.
Risk of bias	Could the selection of participants have introduced bias?	Low risk/high risk/unclear	Summarises previous questions: if any has no as answer, then high risk; if any has unclear, then unclear.
Concerns about applicability	Are there concerns that the included participants and setting do not match the review question?	High concern/Low concern/Unclear	For example, tertiary centres, selected difficult cases. Low concern, if there is a match. Unclear, if there is no such information provided.
Domain 2			
Index test	Were the results of the Forns index and the FIB-4 score interpreted without knowledge of the results of the liver biopsy?	Yes/No/Unclear	Yes for all studies because the test result is automatically calculated.
	If a threshold was used, was it pre-specified?	Yes/No/Unclear	Yes, if the cut-off of Forns index or FIB-4 for a specific fibrosis stage was pre-specified. No, if it was not. Unclear, if there was no such information.
Risk of bias	Could the conduct or interpretation of the Forns index or FIB-4 score have introduced bias?	Low risk/High risk/Unclear	Summarises previous questions: if any has no as answer, then high risk; if any has unclear, then unclear
Concerns about applicability	Are there concerns that the Forns index or FIB-4 score, their conduct, or interpretation differ from the review question?	High concern/Low concern/Unclear	High risk if Forns index or FIB-4 were not conducted according to manufacturer recommendations. Low risk if they are conducted. Unclear if there is no such information.
Domain 3			
Reference standard	Is liver biopsy likely to classify the target condition correctly?	Yes/No/Unclear	Yes, if biopsy length ≥ 6 portal tracts. No if biopsy length < 6 portal tracts. Unclear if there is no such information. Information on length alone insufficient.

(Continued)

	Were the results of liver biopsy interpreted without knowledge of the results of the Forns index or the FIB-4 score?	Yes/No/Unclear	Yes, if the results were interpreted without knowledge of Forns index or FIB-4. No, if they were not interpreted without knowledge of Forns index or FIB-4. Unclear if there is no such information.
Risk of bias	Could liver biopsy, its conduct, or its interpretation have introduced bias?	Low risk/High risk/ Unclear	Summarises previous questions: if any has no as answer, then high risk; if any has unclear, then unclear
Concerns about applicability	Are there concerns that the target condition as defined by liver biopsy does not match the question?	High concern/Low concern/Unclear	Always low concern.
Domain 4			
Flow and timing	Was there an appropriate interval between Forns index or FIB-4 and liver biopsy?	Yes/No/Unclear	Yes, if the interval between biopsy and Forns index or FIB-4 < 3 months. No, if the interval was between 3 and 6 months. Studies with intervals > 6 months were excluded. Unclear if there is no such information.
	Did all participants have liver biopsy?	Yes/No/Unclear	Yes, if all participants had a liver biopsy. No, if only a subset of participants had a liver biopsy. Unclear, if there is no such information.
	Were all participants included in the analysis?	Yes/No/Unclear	Yes, if all participants were included in the analysis. No, if participants with uninterpretable results were not included in the analysis or if there were participants with indeterminate results. Unclear, if there is no such information.
Risk of bias	Could the participant flow have introduced bias?	Low risk/High risk/ Unclear	Summarises previous questions: if any has no as the answer, then high risk; if any has unclear, then unclear risk

HISTORY

Protocol first published: Issue 11, 2015

CONTRIBUTIONS OF AUTHORS

Marc Huttman identified and obtained relevant articles and acquired data from them. MH conducted data analysis and interpretation. MH was the primary author and drafted, revised, and approved the final manuscript.

Tomasso Parigi identified and obtained relevant articles and acquired data from them. TP conducted data analysis and interpretation. TP was a co-author and drafted the manuscript.

Mirko Zoncape identified and obtained relevant articles and acquired data from them.

Antonio Liguori identified and obtained relevant articles and acquired data from them.

Maria Kalafateli drafted the study protocol and identified and obtained relevant articles.

Anna Noel-Storr created the search strategies for this review.

Giovanni Casazza was the senior statistician involved in planning, performing and interpreting the statistical analyses. GC was a co-author and drafted, revised, and approved the final manuscript.

Emmanuel Tsochatzis supervised all stages of the review. ET was the corresponding author and drafted, revised, and approved the final manuscript.

DECLARATIONS OF INTEREST

No conflicts of interest to declare

SOURCES OF SUPPORT

Internal sources

- No sources of support provided

External sources

- The Cochrane Hepato-Biliary Group, Copenhagen Trial Unit, Centre for Clinical Intervention Research, Rigshospitalet, Copenhagen University Hospital, Denmark

  Editorial processes.

DIFFERENCES BETWEEN PROTOCOL AND REVIEW

We changed the title from 'Forns index and 'FIB4' for staging of fibrosis in adults with chronic hepatitis C' to 'Liver fibrosis stage based on the four factors (FIB-4) score or Forns index in adults with chronic hepatitis C'. We made this change to better highlight the target condition and relative dominance of the index tests, and because it was preferred by the reviewing editorial team.

There were some deviations from the published protocol (Kalafateli 2015), described below. Overall, we consider them to represent minor changes to the secondary objectives of the review.

Review criteria. In our protocol, we specified that we would exclude diagnostic case-control studies. However, we decided to include this study design for a more complete picture of the literature, and to rate these studies as high risk of bias for participant selection. We included two diagnostic case-control studies in this review, which contributed a total of 36 participants to one of the meta-analyses. Their inclusion is likely to be inconsequential to the overall results.

In our protocol, a secondary objective was to "compare the diagnostic accuracy of the pre-defined cut-off values, used in the original publications, versus the accuracy of other published cut-off values, used in the included studies". However, in the review, we made more comparisons than this sentence describes, as we felt these comparisons were clinically important, given the variation in cut-offs applied in the included studies.

We had planned to include F0 (no fibrosis) as a target condition. However, it became apparent after the publication of the protocol that the FIB-4 score and the Forns index are not used in clinical practice for this stage. Moreover, the diagnosis of F1 fibrosis is not as clinically significant as more severe fibrosis stages.

Search methods. We had planned to contact research groups who have published or are conducting work on the index test(s) for non-invasive assessment of fibrosis with the initial results of our literature search in order to try and include more data. However, we did not contact research groups asking for new data as we felt our literature search was rigorous and thorough and therefore captured all available data. We did contact authors of papers where there were missing data in otherwise eligible studies.

Statistical analysis and data synthesis. We only included studies in the meta-analysis if the reported cut-offs were in a narrow range around the original validated cut-offs (+/- 0.15 for FIB-4, +/- 0.3 for the Forns index). We did not prespecify this approach in the protocol but decided to adopt it for the review in order to produce more robust summary estimates with lower heterogeneity, and to provide meaningful results for clinical practice. This narrow range approach is not a multiple threshold approach but an approach to examine the diagnostic accuracy of a defined threshold with small non-significant deviations. This is in line with narrow ranges in the cut-offs of non-invasive tests in the recent EASL guidelines (EASL 2021).

Investigations of heterogeneity. We were not able to explore histological inflammation levels as a potential source of heterogeneity as outlined in the protocol, due to scant data. We were only able to investigate the quality of liver biopsy as a potential source of heterogeneity in two index test/cut-off combinations. Similarly, due to scant data, we were not able to compare the diagnostic accuracy of the Forns index and the FIB-4 score across all fibrosis stages by using the diagnostic tests as a covariate in the bivariate model. We were only able to compare low cut-off FIB-4 versus low cut-off Forns index for significant fibrosis and higher (≥ F2) and high cut-off FIB-4 versus high cut-off Forns index for significant fibrosis and higher (≥ F2).

Although the FIB-4 score was initially developed to evaluate severe fibrosis, and the Forns index for significant fibrosis, we expanded the clinical setting in our review to include other settings (e.g. FIB-4 to assess significant fibrosis) because we found several studies evaluating these settings (e.g. using FIB-4 for the assessment of ≥ F2) and wanted to further explore if this was clinically meaningful.

We did not create a graph of pre-test probabilities, but did report them in the main section of results.

Sensitivity analysis. We had planned to perform a sensitivity analysis by excluding studies with a high risk of bias; that is, those with a high risk of bias rating in one or more domains. We instead performed a sensitivity analysis by excluding studies with either a high or unclear risk of bias. In addition to what was described in the protocol, we conducted a sensitivity analysis by excluding studies published as abstracts only.