Accepted Article

DR. EMMANUEL A TSOCHATZIS (Orcid ID : 0000-0001-5069-2461)

Article type : Original

Defining the minimum acceptable diagnostic accuracy of non-invasive fibrosis testing in cirrhosis: a decision analytic modelling study

Short Title: Acceptable diagnostic accuracy of non-invasive fibrosis testing in cirrhosis: a decision analytic modelling study

Authors:
Avik Majumdar[1,2,3,4] (avik.majumdar@health.nsw.gov.au)
Sara Campos[1,2,5] (saratcampos@gmail.com)
Kurinchi Gurusamy[6] (k.gurusamy@ucl.ac.uk)
Massimo Pinzani[1,2] (m.pinzani@ucl.ac.uk)
Emmanuel A. Tsochatzis[1,2] (e.tsochatzis@ucl.ac.uk)

1. UCL Institute for Liver and Digestive Health, Royal Free Hospital and UCL, London, UK
2. Sheila Sherlock Liver Centre, Royal Free Hospital, London, UK
3. AW Morrow Gastroenterology and Liver Centre, Royal Prince Alfred Hospital, Sydney, Australia
4. Department of Gastroenterology, The Alfred Hospital, Melbourne, Australia
5. Department of Gastroenterology, Centro Hospitalar e Universitário de Coimbra, Portugal
6. Department of Surgery, Royal Free Campus, UCL Medical School, London, UK

Keywords: Net benefit, fibrosis, screening, Fibroscan, APRI, FIB4, Fibrotest

Contact Information:

Dr Emmanuel A. Tsochatzis
Royal Free Hospital
Pond St, Hampstead
London NW32QG
United Kingdom
Email: e.tsochatzis@ucl.ac.uk
Tel: +44 2074332784

Dr Avik Majumdar
Royal Prince Alfred Hospital
Missenden Rd, Camperdown
NSW 2050
Australia
Email: avik.majumdar@health.nsw.gov.au
Tel: +61 2 9515 8578

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/hep.30846
This article is protected by copyright. All rights reserved.

List of Abbreviations: NIT, non-invasive test; LB, liver biopsy; P_t, threshold probability; VCTE, vibration controlled transient elastography; HCC, hepatocellular carcinoma; AUROC, area under the receiver operating characteristic curve; FN, false negative; FP, false positive; TN, true negative; TP, true positive; DSA, deterministic sensitivity analysis; PSA, probabilistic sensitivity analysis; APRI, AST to platelet ratio index; HS, high specificity; LS, low specificity; mNIT, minimum sensitivity and specificity for equivalent mortality between liver biopsy and non-invasive testing; AGA, American Gastroenterological Association.

Author Contributions:
AM – manuscript preparation, data acquisition, statistical analysis, study concept
SC – data acquisition and critical review of manuscript
KG –wrote the computer codes necessary for the analysis with Excel, data acquisition, methodological advice, and critical review of manuscript, study concept
MP – interpretation of data, critical review of manuscript, study concept
ET – data acquisition, interpretation of data, critical review of manuscript, study concept and design

Financial Support: Nil

Abstract

No studies explore the clinical consequences of using non-invasive tests (NITs) compared to liver biopsy (LB) in diagnosing cirrhosis. Our aim was to combine two decision analytic models to determine the minimum diagnostic accuracy criteria for NITs to diagnose cirrhosis with equivalence to LB in terms of mortality. We further evaluated selected existing NITs used alone and sequentially. A decision tree was constructed with associated 2-year mortality incorporating a LB or NIT strategy to diagnose cirrhosis in a hypothetical cohort of 1000 asymptomatic patients. Cirrhosis prevalence was modelled at 5%, 20% and 50%. Decision curve analyses were performed, expressing the net benefit of tests over a range of threshold probabilities (P_t). The NIT deriving from the two models that could diagnose cirrhosis with at least equal mortality to LB was termed mNIT. Existing NITs were then compared using both decision models. The combined mNIT minimum sensitivity and specificity to diagnose cirrhosis with equivalence to LB at 5%, 20% and 50% cirrhosis prevalence were; 89% and 88%, 94% and 85%, and 94% and 87%, respectively at P_t=0.20. Sequential NITs performed better than single NITs at any prevalence. Combining both decision models, FibroTest® plus VCTE (vibration controlled transient elastography) and VCTE alone were the only existing NITs that were better or equal to LB at diagnosing cirrhosis at 5% prevalence. At 20% and 50% prevalence, only FibroTest® high specificity cut-off plus VCTE was equivalent or better than LB.

This article is protected by copyright. All rights reserved.

Accepted Article

Conclusion: Decision analytic models were used to determine the minimum acceptable diagnostic accuracy of NITs for diagnosing cirrhosis. We recommend that such models should be used as the new standard in evaluating the diagnostic performance of NITs.

Introduction

Non-invasive tests (NITs) for diagnosing cirrhosis have rapidly progressed from development to acceptance in consensus guidelines.(1, 2) The basis for this progression is the vast amount of published data comparing specific NITs to liver biopsy (LB). NITs have generally been developed to detect the presence of either significant fibrosis or compensated cirrhosis prior to the development of overt clinical manifestations. The importance of diagnosing cirrhosis at a pre-clinical stage is to identify patients at risk of complications. Specifically, cirrhosis signifies the point at which to commence surveillance for hepatocellular carcinoma (HCC) and gastroesophageal varices, as well as aggressive aetiology-specific management if not already instituted.(3)

No minimal performance criteria for NITs currently exist to govern their clinical use. The performance of NITs have generally been expressed in the literature as areas under the receiver operating characteristic curve (AUROCs), sensitivity and specificity. However, such statistical metrics of accuracy give no information on whether a test is fit for clinical practice.(4) The clinical implications of an NIT that results in a missed diagnosis (false negative, FN) or conversely, misclassifying a patient as having cirrhosis when they do not (false positive, FP), are not taken into account. Furthermore, an acceptable "trade-off" between missed or over-diagnosed cases of cirrhosis cannot be determined using the above accuracy metrics alone.(5)

Decision analytic models seek to address these issues through quantitative evaluation of the consequences of using specific diagnostic strategies. Two major methods of decision analysis have been described in this context; decision trees and decision curves. Decision tree models determine the individual probabilities associated with each possible outcome of diagnostic strategies based on the published literature. Decision curves quantify the net benefit of using diagnostic strategies over a range of threshold probabilities (P_t).(6, 7) Each method has its

This article is protected by copyright. All rights reserved.

Accepted Article

limitations, with the former being dependent on the quality of the literature for outcome probabilities and the latter involving P_t, which may vary amongst individual clinicians and patients. We hypothesised that by combining and comparing the two types of analysis, the uncertainty in the benefits of a diagnostic test can be estimated.

The aims of this study were to determine the minimal diagnostic criteria for NITs in detecting compensated cirrhosis at an equivalent associated mortality to LB and to evaluate the clinical utility of existing NITs in this context, on a single-test (one-off test) or sequential-test (using a second NIT depending on the results of the previous test) basis.

$$Net\ benefit = \frac{TP}{N} - \frac{FP}{N} \times \frac{P_t}{1 - P_t}$$

Materials and Methods

NITs and LB were evaluated in three scenarios that varied the prevalence of cirrhosis at 5%, 20% (base case) and 50% to represent populations at risk in primary care (5%), and secondary or tertiary care settings (20% and 50%), respectively. The base case prevalence was derived from a recent systematic review and meta-analysis funded by the UK NIHR-HTA program that investigated the cost-effectiveness of non-invasive assessment methods of assessing liver fibrosis, where the average prevalence of cirrhosis for all aetiologies in secondary care was 20%. The 5% prevalence refers to a population at increased risk of cirrhosis in primary care that could potentially benefit from screening and was derived from the base case analysis of cirrhosis in unselected patients with NAFLD in the same project.(8) The 50% prevalence refers to a high clinical suspicion of cirrhosis after additional tests in secondary care.

The baseline sensitivity and specificity of LB for diagnosing cirrhosis was evaluated at 95% and 100% respectively (no FPs).

Decision tree

A decision tree was constructed to determine the probability of death using a LB strategy or a NIT strategy to diagnose compensated cirrhosis (Figure 1). A sample size calculation was performed to include all three prevalence scenarios and

This article is protected by copyright. All rights reserved.

Accepted Article

variations in LB sensitivity(9). A hypothetical cohort of 1,000 patients was examined. The decision tree pathways involved the consequences of a true positive (TP), false positive (FP), true negative (TN) or false negative (FN) diagnosis of cirrhosis with NIT versus LB. Those classified as having cirrhosis (TPs or FPs) underwent ultrasound screening for HCC and endoscopic variceal surveillance. A FN diagnosis of cirrhosis resulted in an increased possibility of death due to HCC or bleeding from missed varices. A FP diagnosis of cirrhosis led to the outcome of unnecessary endoscopy and ultrasound surveillance. A true negative (TN) result had no additional consequences beyond the test used (LB or NIT).

The minimal acceptable diagnostic accuracy criteria for NITs were determined by finding the sensitivity and specificity values where the mortality associated with the NIT strategy was equivalent to LB and the test that fulfilled these criteria was termed mNIT. The performance of existing NITs was then evaluated in the model by inputting their sensitivity and specificity data.(8) The decision tree was programmed with Microsoft® Visual Basic for Applications in Microsoft Excel 2010®.

The robustness of the decision tree model was initially tested with multiple one-way deterministic sensitivity analyses (DSA). Input variables were varied within pre-determined ranges based on literature review. The variable ranges were derived from the 95% confidence intervals from the input data where available. In cases where the input data were based on observational studies (e.g. mortality associated with upper gastrointestinal endoscopy), the range was determined from similar sized observational studies where available, or published expert consensus opinion otherwise. DSA results were expressed as a tornado diagram using a base case LB sensitivity of 95% and cirrhosis prevalence of 20%. Probabilistic sensitivity analysis (PSA) was also used. PSA involves assigning a range of distributions to all input variables and then using Monte Carlo simulation to sample from these distributions. Beta-distributions were applied to all input variables and 10,000 iterations were performed.(10) TPs, FPs, TNs and FNs were expressed as a percentage of the total hypothetical cohort of 1000 patients.(11)

This article is protected by copyright. All rights reserved.

Accepted Article

Decision curve analysis and net benefit

As the mortality risk of an endoscopy for variceal surveillance is negligible (1:40,000) and there is no mortality risk from unnecessarily undergoing surveillance ultrasonography, a meaningful difference in mortality between patients with TN and FP diagnoses of cirrhosis would not be evident in the decision tree model. However, a FP diagnosis carries the impacts of unnecessarily undergoing invasive endoscopies, hospital visits, further testing following FP screening ultrasound results and the psychosocial burden of the diagnosis of cirrhosis. To further investigate the utility of NIT strategies and account for the wider impacts of a FP diagnosis of cirrhosis, we used decision curve analysis as a complementary method to the decision tree model.

Decision curve analysis involves calculating the net benefit, which directly compares the harms and benefits of a test by multiplying the harm by a threshold probability so that it is placed on the same scale as benefit.(7) Net benefit can be quantified by using the following formula:

The clinical interpretation of net benefit can be explained by comparison to a "treat none" strategy. For example, if the net benefit value of an NIT was 0.16, the NIT is the equivalent of a test that correctly diagnoses and treats 16 per 100 patients with cirrhosis without treating any patients unnecessarily, compared to no patients being treated as cirrhotic. Threshold probability is the chance of a correct diagnosis that a clinician or a patient is willing to accept to undertake a test. For example, if a clinician found it acceptable to subject ten patients to a LB to find one correct diagnosis of cirrhosis, then the P_t is 10%. The methods of Vickers and Elkin(6, 7) were used to calculate the net benefit over a range of P_t and were expressed as decision curves. The range of acceptable P_t was 0.2 to 0.3 (3 to 5 patients biopsied for one correct diagnosis of cirrhosis), with 0.2 being the most common response from structured questions to ten senior hepatologists and five patients with chronic liver disease (Supplementary Table 1).

LB and NIT diagnostic strategies were compared to the reference curves of "Treat All" (treat all as having cirrhosis) and "Treat None" (treat none as having cirrhosis resulting in no net benefit). The reference curves intersect at the chosen cirrhosis

This article is protected by copyright. All rights reserved.

Accepted Article

prevalence. LB sensitivity and specificity were modelled at 95% and 100%, respectively. This means that LB resulted in no FPs and therefore, the net benefit does not change with P_t (indicated by the net benefit formula) and a flat curve is produced.

A factor of harm of 3.2% was applied to LB. This was necessary in the decision curve model to account for the impacts of an invasive test, compared to a NIT. Harm, as it applies to invasive tests, is a holistic estimate of the negative consequences of the test including medical harms, financial cost, loss of productivity, psychological stress and inconvenience amongst other factors. The harm is quantified by determining how many tests could be justifiably performed for a single correct diagnosis, assuming the test was perfect. The 3.2% harm added to all analyses is the equivalent of 31 biopsies performed safely for a single diagnosis of cirrhosis. This was based on a historic hospitalisation rate for complications of outpatient LB (12) and resulted in a net benefit reduction of 0.032 for LB across all P_t. This factor of harm was selected to represent the aforementioned holistic impacts of LB and was guided by the results of the structured questionnaire (Supplementary Table 1). We assumed that although ultrasound guided LB advances may have reduced the hospitalization rate over time, with more recent estimates of major adverse events/hospitalization varying between 1.7%-3.5%(13), a harm of 3.2% would allow some concession for the additional impacts of LB. Omitting the harm factor entirely from LB would result in NITs only being equivalent to liver biopsy at clinically irrelevant threshold probabilities (P_t=0.02 or less, 50 or more LBs for one cirrhosis diagnosis). An explanation of the impact of harm for invasive tests on decision curves is presented in the Supplementary Materials. All net benefit and decision curve calculations were performed using STATA software version 15 (STATACorp®, USA).

Agreement between decision analytic models

The mortality from a FN diagnosis was quantified by the decision tree analysis and hence this was used to derive the sensitivity of mNIT. The decision curve analysis quantified the trade-off between FP and TP diagnoses and therefore was more appropriate to determine the mNIT specificity. The specificity was adjusted so that mNIT had the equivalent net benefit as LB at a P_t of 0.20 (Supplementary Figure 1),

This article is protected by copyright. All rights reserved.

Accepted Article

which was the most common acceptable P_t from our survey of patients and clinicians, to derive the combined mNIT. Figure 2 summarises the methods for the combined mNIT. Existing NITs demonstrated agreement between the two decision analytic models if the NITs had mortality rates equal or less than LB on decision tree analysis; and had greater net benefit than the combined mNIT on decision curve analysis.

Input data

Input data for the probabilities of death due to each outcome in the decision tree model are summarised in Table 1. Studies were identified through structured searches of Medline and PubMed using combinations of the following search terms; “cirrhosis”, “mortality”, “death”, “oesophageal varices”, “hepatocellular carcinoma”, “surveillance”, “screening”, “liver biopsy”, “varices”, “endoscopy” and “gastroscopy”. Where possible, randomised clinical trial data were used. Where no data were available, expert opinion was sought. The time horizon for the decision tree model was set at 2 years as the mortality data in existing systematic reviews of randomised controlled trials were available for a follow-up of 2 years.

The pooled sensitivity and specificity for the four most-studied NITs for diagnosing cirrhosis of any aetiology were used; Vibration Controlled Transient Elastography (VCTE), FibroTest®, AST to platelet ratio index (APRI) and FIB-4.(8) The three serum NITs (FIB-4, APRI and FibroTest®) have dual cut-off values that stratify patients into high risk, low-risk and indeterminate risk groups.(8) High specificity (HS) and low specificity (LS) cut-offs were evaluated separately (Supplementary Table 2).

The sequential use of NITs was modelled on real-world scenarios (Figure 3). Simple, non-proprietary scores such as FIB-4 and APRI were used as first-line tests, with VCTE reserved as a second-line test. The combination of Fibrotest® followed by VCTE was also investigated. HS, LS and Dual (both HS and LS) cut-offs were used for all first-line tests. We assumed that all testing strategies could be successfully applied to the study cohort.

This article is protected by copyright. All rights reserved.

Accepted Article

Sensitivity Analyses

We performed the following sensitivity analyses:

1) The LB sensitivity was increased to 100%. Serum markers of fibrosis have been developed and calibrated to individual sets of reference liver biopsies. The ideal serum marker could theoretically achieve perfect sensitivity and specificity when compared to LB, replicating even the misclassification of the reference biopsy set.(14)
2) The LB sensitivity was reduced to 80% to account for sampling variability, particularly in macronodular cirrhosis.(15-17)
3) The mortality from cirrhosis was reduced by 10% and by 50% over the two-year horizon to account for the impact of aetiology specific treatment such as viral eradication in chronic hepatitis C, abstinence in alcohol related liver disease or weight loss in non-alcoholic fatty liver disease.
4) The prevalence of cirrhosis was modelled at 1% as population estimates of cirrhosis prevalence range from 0.1% to 2.1%.(18, 19)
5) The probability of varices when cirrhosis was present was reduced to 34.8% to reflect more contemporaneous estimates from observational studies.(20, 21)
6) The mortality from HCC in surveilled and unsurveilled patients was altered using data from a recent meta-analysis of observational studies.(22)
7) To account for potential improvements in the performance of existing NITs and changes in the casemix of liver disease aetiologies since the publication of the input data(8), the upper and lower confidence interval limits for the summary sensitivity and specificity for existing NITs were examined.
8) The harm associated with the invasiveness of LB was increased to 5%, which is the equivalent of 20 biopsies for a single diagnosis of cirrhosis.

Accepted Article

This article is protected by copyright. All rights reserved.

Accepted Article

Results

Decision Tree mNIT and Single NITs

The results of the decision tree model using a baseline LB sensitivity of 95% are summarised in Table 2. The 2-year mortality associated with using LB to diagnose cirrhosis at 5%, 20% and 50% cirrhosis prevalence was 0.41%, 1.54% and 3.81%, respectively. Consequently, the decision tree mNIT sensitivity at 5%, 20% and 50% prevalence was 89%, 94% and 94% with corresponding specificity of 75%, 63% and 87%, respectively. At a prevalence of 5%, VCTE and FibroTest® LS were the only single NITs that resulted in equal or lower mortality than LB. Compared to FibroTest® LS, VCTE had lower FNs, lower FPs and a greater percentage of patients correctly classified. No single NITs outperformed LB at 20% or 50% prevalence.

Decision Tree Sequential NITs

The HS cut-offs for serum NITs resulted in the most favourable sensitivity, specificity and mortality rate. At 5% and 20% cirrhosis, the HS cut-offs for APRI, FIB-4 or FibroTest® followed by VCTE (Figure 3B) all had lower mortality than LB. Of these test strategies, FibroTest® HS plus VCTE had the lowest mortality. Only FibroTest® HS plus VCTE performed better than LB at 50% prevalence (Table 2).

Decision Tree Sensitivity Analyses

Multiple one-way DSAs are summarised in Supplementary Figure 2. Altering the prevalence of cirrhosis had the greatest influence on overall mortality in the model. As anticipated, varying NIT specificity and hence the number of FPs had little effect on mortality, whereas the opposite was true for NIT sensitivity. Varying the probability of death from a missed diagnosis of HCC also did not greatly influence overall mortality in the model, particularly as this variable was dependent on both the prevalence of cirrhosis and the probability of developing HCC in the first instance. PSA demonstrated no difference in mean mortality between the decision tree mNIT and LB (Supplementary Figure 3).

Additional analyses varying LB sensitivity and prevalence are presented in Supplementary Tables 3 and 4, respectively. By reducing two-year mortality by 10% and 50% due to aetiological treatments for cirrhosis, the specificity of the combined

This article is protected by copyright. All rights reserved.

mNIT could be decreased slightly at 50% cirrhosis prevalence, but not at lower prevalence (Supplementary Tables 5A, 5B). Reducing the probability of varices when cirrhosis was present lowered the associated mortality of the model, but did not affect the mNIT (Supplementary Table 6). Using input data from a recent meta-analysis of observational studies increased the mortality associated with HCC in the decision tree model (Supplementary Table 7). This was also the case when only considering data from studies including less than 10% Child-Pugh B and C patients (Supplementary Table 8).

Decision Curve Analysis

The best performing test strategies (NITs vs. LB) according to greatest net benefit over various P_t, independent of the outcome of the decision tree model, are summarised in Table 3. Supplementary Table 9 demonstrates the P_t above which an NIT based strategy provides greater net benefit than LB.

At 5% prevalence, the single NITs with the highest net benefit were FibroTest® HS and VCTE. FibroTest® LS had lower net benefit than decision tree mNIT in the decision curve analysis despite having equivalent mortality to the decision tree mNIT (Supplementary Figure 4). This is explained by the high FPs for FibroTest® LS compared to the decision tree mNIT (33.2/100 vs. 23.8/100). Sequential test combinations had greater net-benefit than single NITs (Table 3).

At 20% prevalence, VCTE was the only single NIT with greater net-benefit than LB up to a P_t of 0.18. FibroTest® Dual cut-off plus VCTE gave similar net-benefit to VCTE alone. The sequential use of FibroTest® HS plus VCTE was the best test strategy up to P_t=0.21, after which LB was superior.

No single existing NITs, excluding decision tree mNIT, were superior to LB at 50% prevalence. The sequential use of FibroTest® HS plus VCTE was the best test strategy for P_t up to 0.33. The decision tree mNIT provided greater net benefit than LB until a P_t of 0.30. This is the equivalent of a LB strategy where three patients are subjected to biopsies for a single correct diagnosis of cirrhosis.

This article is protected by copyright. All rights reserved.

Accepted Article

Agreement between decision analytic models

The combined mNIT sensitivity and specificity derived from both decision models at 5%, 20% and 50% cirrhosis prevalence were; 89% and 88%, 94% and 85%, and 94% and 87%, respectively. Table 4 and Figure 4 summarise the NITs that demonstrated agreement between models. Of the tests that agreed at 5% prevalence, the sequential use of FibroTest® Dual plus VCTE (sensitivity 87%, specificity 91%) provided the highest net benefit between P_t 0.09-0.24. Above P_t 0.24 (<4.2 biopsies for one correct diagnosis of cirrhosis), LB was the most effective strategy (Figure 4A). VCTE was the only single NIT that was fit for practice from the decision tree model and was the next best test strategy on decision curve analysis. Although the FibroTest® Dual cut-off plus VCTE strategy carried higher overall mortality in the decision tree than VCTE alone (0.41% vs. 0.40%), the net benefit was greater. This suggests that the increased FNs from 0.5/100 to 0.7/100 are outweighed by the reduction in FPs from 10.4/100 to 8.7/100 in this strategy.

At 20% and 50% prevalence, only the sequential strategy of FibroTest® HS plus VCTE (sensitivity 97%, specificity 84%) demonstrated agreement between models. This strategy had the greatest net benefit at P_t of up to 0.21 and 0.33 at 20% and 50% prevalence, respectively. If either one of the sensitivity or the specificity of the combined mNIT was exceeded by an existing NIT strategy to provide higher net-benefit, then the other parameter (sensitivity or specificity) could be marginally lower than the combined mNIT.

The sensitivity analysis using the upper confidence interval limits from the Crossan meta-analysis for existing NITs resulted in more NITs demonstrating agreement between the two models (Supplementary Table 10). All existing NITs had higher mortality than LB when using the lower confidence interval limits.(8)

Finally, increasing the factor of harm of LB to 5% resulted in additional NITs that demonstrated agreement between the two models, but at the cost of more FP diagnoses. At cirrhosis prevalence of 5%, LB had a lower net benefit than the treat none strategy. This is an implausible clinical scenario and the adjusted combined mNIT at P_t=0.20 (specificity 85%) resulted in zero net benefit in this situation. Using this combined mNIT, all sequential NITs using the HS cut-off plus VCTE

This article is protected by copyright. All rights reserved.

Accepted Article

demonstrated agreement between the two models compared to the baseline analysis (Supplementary Figure 5A). At 20% prevalence (Supplementary Figure 5B), all sequential NITs using the HS cut-off plus VCTE also demonstrated agreement, as the combined mNIT specificity was reduced to 75%. At 50% cirrhosis prevalence, no NITs demonstrated agreement compared to the baseline analysis (Supplementary Figure 5C).

Discussion

Key findings

We investigated the role of NITs in diagnosing cirrhosis compared to LB using two robust methods of decision analytic modelling across a range of clinical scenarios. We constructed a decision tree model that quantified the mortality associated with outcomes of NIT strategies and we used decision curve analysis to account for the impact of FP diagnoses. Using the combination of these models, we established minimum sensitivity and specificity values (mNIT) for a NIT strategy to be equivalent in mortality outcomes to a LB strategy for low (5%), medium (20%) and high (50%) cirrhosis prevalence settings. This work therefore provides the benchmark for adopting NITs in clinical practice. We further assessed existing NIT strategies and their fitness for practice. Using both decision analytic models, we found that the sequential use of Fibrotest® Dual plus VCTE provides the greatest diagnostic utility at 5% prevalence followed by VCTE alone. Both FIB-4 HS plus VCTE or APRI HS plus VCTE had lower mortality than LB at this prevalence, however the high number of FPs resulted in lower net benefit. In practical terms, these strategies may be used in primary care as screening tools in patients at risk and guide referrals to secondary care, as fewer cirrhotic patients would be missed than VCTE alone, however the additional FPs would have to be reconciled. At higher prevalence of cirrhosis, the sequential use of FibroTest® HS plus VCTE was the only suitable strategy and can be used instead of a liver biopsy with similar clinical outcomes. This suggests that in people with indeterminate or low-risk FibroTest® results, a negative VCTE is required as a second line test to safely exclude cirrhosis. FibroTest®, APRI and FIB-4 are not suitable standalone tests for diagnosing cirrhosis at any prevalence.

This article is protected by copyright. All rights reserved.

Accepted Article

Relationship to previous studies

To our knowledge, this is the first study where the clinical utility of NITs in cirrhosis was investigated beyond traditional accuracy metrics and cost-effective analysis. Decision tree models have previously been used extensively to investigate NITs for the purposes of cost-effectiveness, but uncommonly for the endpoint of mortality.(8, 23, 24) Similarly, decision curves and net benefit have been used to assess the performance of a broad range of diagnostic strategies, such as the QRISK2 cardiovascular risk score and prostate cancer algorithms, but their use in the non-invasive assessment of liver fibrosis is lacking.(25, 26) Other forms of decision analysis such as "relative utility" and "weighted net reclassification improvement" have been shown to achieve similar results to net benefit.(27)

A recent American Gastroenterological Association (AGA) technical review suggested that it was acceptable if less than 5-10% of patients had a missed diagnosis of cirrhosis (FN) when choosing cut-off values of VCTE to diagnose cirrhosis.(11) This threshold was established from expert opinion and did not involve any dedicated decision analysis. No acceptable threshold for FP diagnoses of cirrhosis using VCTE was specified in the AGA review, however, FPs varied from to 6.3/100 to 27.5/100 depending on the aetiology and prevalence of cirrhosis. The acceptable number of FPs in our study was determined by decision curve analysis and hence is dependent on a probability threshold being chosen by a clinician and/or patient. For example, assuming an acceptable risk of 5 biopsies or less for each correct diagnosis of cirrhosis ($P_t \geq 0.20$), our data suggest that the equivalent FPs for a NIT are less than or equal to 11.4/100, 12.0/100 and 6.5/100 at 5%, 20% and 50% prevalence, respectively. We concur with the AGA review that VCTE is superior to FIB-4 and APRI alone.

Our findings also support the observation that the sequential use of NITs is superior to single tests in the diagnosis of cirrhosis, under the assumption that the diagnostic accuracy of the second tier NIT is similar when used singly or sequentially.(1, 28-31)

This article is protected by copyright. All rights reserved.

Accepted Article

Implications of study findings

Our study moves beyond the limitations of traditional statistical metrics in assessing the performance of NITs and considers the outcome of a testing strategy. Clearly, the mortality risks of a FN diagnosis of cirrhosis are far greater than a FP diagnosis. However, the implications of a FP diagnosis are more difficult to quantify objectively and include psychosocial, financial and opportunity costs. We addressed these issues by combining decision analytic models and looking for agreement between the results of each model. The decision tree model effectively quantified the mortality associated with test outcomes based on systematic literature reviews. The decision curve analysis quantified the trade-off between FP and true positive diagnoses at threshold probabilities that a patient and/or clinician determine as acceptable.

Both types of decision analysis have been used in diverse settings and hence are generalizable to other diagnostic strategies that require comparison. We therefore propose that decision analytic models become the new standard in reporting and comparing the performance of NITs.

Strengths and Limitations

Our study has several strengths. The methodological concepts used are robust and their application in diagnosing cirrhosis is unique. We have used input data for our models from structured literature searches and using pooled NIT data from the largest published systematic review of NITs for cirrhosis. The use of pooled meta-analysis data has reduced the influence of spectrum bias that is often observed in single-centre studies of NITs at large liver referral centres. We have performed numerous sensitivity analyses and modelled several clinical scenarios. The models are easily adapted to other NITs and therefore our study can provide a framework to guide the real world use of NITs in diagnosing cirrhosis.

There are also limitations to our study. Firstly, the model input data contain variability in the aetiologies of cirrhosis, the populations and eras studied as well as the overall level of evidence and quality of included studies. However, the input data for the decision tree analysis is the result of a thorough, structured literature search. Meta-analysis data that includes patients enrolled in HCC and variceal surveillance(32),

This article is protected by copyright. All rights reserved.

Accepted Article

suggest that the probability of death in compensated cirrhosis is similar to the two-year mortality rates from the decision tree model, where if cirrhosis prevalence was 100%, the associated mortality would be 7.3%.

We were unable to find any higher quality studies and found that separating the analysis by aetiologies of cirrhosis led to further fragmentation of the evidence base. For example, the study by Zhang and colleagues(33) is the only randomised control trial where the probability of HCC in screened versus non-screened patients is investigated, but only includes cirrhotic and non-cirrhotic patients with hepatitis B and has had methodological criticism in the past. Yet, this study forms the basis of consensus recommendations for HCC screening in cirrhosis(34). We used alternative input data from a recent meta-analysis (22), however only 3-year mortality data was available and hence resulted in higher mortality rates in the decision tree model than the Zhang study. This highlights the fact that although the evidence base is not flawless in cirrhosis, we used the best available representation. We addressed variability in the data by performing several additional analyses, including deterministic and probabilistic sensitivity analyses, and adopted a conservative approach when assessing the level of evidence of input data and mortality estimates in the modelling. Furthermore, altering the input variables in the decision tree had little impact on the mNIT, even when mortality was reduced by 50%, validating our findings.

Secondly, by using the results of a previous meta-analysis for NIT data, our analysis is subject to the same limitations, such as the lack of included studies beyond 2012 and the low proportion of studies of high methodological quality. However, the Crossan meta-analysis(8) remains the largest and most detailed of its nature, used similar cirrhosis prevalence as our models and was also the basis of the AGA technical review(11).

Fourthly, we did not use an intention to diagnose approach and therefore were not able to incorporate the failure rate of individual testing in our modelling. We assumed that all included patients had applicable tests.

This article is protected by copyright. All rights reserved.

Accepted Article

Finally, both forms of decision models have inherent weaknesses. The dependence on the published literature is problematic when constructing a decision tree if good quality evidence is not available. Conversely, the dependence of decision curve analysis on threshold probability is not completely objective and requires individual interpretation by clinicians and/or patients. Similar issues exist when considering the harms of LB in the decision curve analyses, which we aimed to mitigate by using the available literature and the results of the structured questionnaire. Importantly, we found that the two decision analytic models chosen were in fact complementary and agreement was demonstrated. The decision tree model better quantified the mortality of FN diagnoses, whilst the decision curve model expressed the relative impact of FP diagnoses.

Conclusions

We have established the minimum acceptable diagnostic accuracy of NITs in diagnosing cirrhosis by using decision analytic models. For existing NITs, their sequential use provides the best diagnostic utility compared to single NIT strategies. The combination of either APRI or FIB-4 followed by VCTE can be used as screening tests in settings of low prevalence of cirrhosis. The combination of FibroTest® HS plus VCTE can be used instead of liver biopsy at higher prevalence. Decision analytic models should be considered as the new standard of care in evaluating and comparing NIT diagnostic strategies.

References

1. EASL-ALEH Clinical Practice Guidelines: Non-invasive tests for evaluation of liver disease severity and prognosis. J Hepatol 2015;63:237-264.

2. de Franchis R, Baveno VIF. Expanding consensus in portal hypertension: Report of the Baveno VI Consensus Workshop: Stratifying risk and individualizing care for portal hypertension. J Hepatol 2015;63:743-752.

3. Tsochatzis EA, Bosch J, Burroughs AK. Liver cirrhosis. Lancet 2014;383:1749-1761.

4. Guha IN, Myers RP, Patel K, Talwalkar JA. Biomarkers of liver fibrosis: what lies beneath the receiver operating characteristic curve? Hepatology 2011;54:1454-1462.

5. Mallett S, Halligan S, Thompson M, Collins GS, Altman DG. Interpreting diagnostic accuracy studies for patient care. BMJ 2012;345:e3999.

This article is protected by copyright. All rights reserved.

Accepted Article

6. Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. Med Decis Making 2006;26:565-574.

7. Vickers AJ, Van Calster B, Steyerberg EW. Net benefit approaches to the evaluation of prediction models, molecular markers, and diagnostic tests. BMJ 2016;352:i6.

8. Crossan C, Tsochatzis EA, Longworth L, Gurusamy K, Davidson B, Rodriguez-Peralvarez M, Mantzoukis K, 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 Technol Assess 2015;19:1-409, v-vi.

9. Flahault A, Cadilhac M, Thomas G. Sample size calculation should be performed for design accuracy in diagnostic test studies. J Clin Epidemiol 2005;58:859-862.

10. Claxton K, Sculpher M, McCabe C, Briggs A, Akehurst R, Buxton M, Brazier J, et al. Probabilistic sensitivity analysis for NICE technology assessment: not an optional extra. Health Econ 2005;14:339-347.

11. Singh S, Muir AJ, Dieterich DT, Falck-Ytter YT. American Gastroenterological Association Institute Technical Review on the Role of Elastography in Chronic Liver Diseases. Gastroenterology 2017;152:1544-1577.

12. Janes CH, Lindor KD. Outcome of patients hospitalized for complications after outpatient liver biopsy. Ann Intern Med 1993;118:96-98.

13. Maheux A, Purcell Y, Harguem S, Vilgrain V, Ronot M. Targeted and non-targeted liver biopsies carry the same risk of complication. Eur Radiol 2019.

14. Tsochatzis EA, Germani G, Hall A, Anousou PM, Dhillon AP, Burroughs AK. Noninvasive assessment of liver fibrosis: the need for better validation. Hepatology 2011;53:1781-1782; author reply 1782-1783.

15. Abdi W, Millan JC, Mezey E. Sampling variability on percutaneous liver biopsy. Arch Intern Med 1979;139:667-669.

16. Regev A, Berho M, Jeffers LJ, Milikowski C, Molina EG, Pyrsopoulos NT, Feng ZZ, et al. Sampling error and intraobserver variation in liver biopsy in patients with chronic HCV infection. Am J Gastroenterol 2002;97:2614-2618.

17. Bravo AA, Sheth SG, Chopra S. Liver biopsy. N Engl J Med 2001;344:495-500.

18. Fleming KM, Aithal GP, Solaymani-Dodaran M, Card TR, West J. Incidence and prevalence of cirrhosis in the United Kingdom, 1992-2001: a general population-based study. J Hepatol 2008;49:732-738.

19. Scaglione S, Kliethermes S, Cao G, Shoham D, Durazo R, Luke A, Volk ML. The Epidemiology of Cirrhosis in the United States: A Population-based Study. J Clin Gastroenterol 2015;49:690-696.

This article is protected by copyright. All rights reserved.

Accepted Article

20. Augustin S, Pons M, Maurice JB, Bureau C, Stefanescu H, Ney M, Blasco H, et al. Expanding the Baveno VI criteria for the screening of varices in patients with compensated advanced chronic liver disease. Hepatology 2017;66:1980-1988.

21. D'Amico G, Pasta L, Morabito A, D'Amico M, Caltagirone M, Malizia G, Tine F, et al. Competing risks and prognostic stages of cirrhosis: a 25-year inception cohort study of 494 patients. Aliment Pharmacol Ther 2014;39:1180-1193.

22. Singal AG, Pillai A, Tiro J. Early detection, curative treatment, and survival rates for hepatocellular carcinoma surveillance in patients with cirrhosis: a meta-analysis. PLoS Med 2014;11:e1001624.

23. Tsochatzis EA, Crossan C, Longworth L, Gurusamy K, Rodriguez-Peralvarez M, Mantzoukis K, O'Brien J, et al. Cost-effectiveness of noninvasive liver fibrosis tests for treatment decisions in patients with chronic hepatitis C. Hepatology 2014;60:832-843.

24. Gaidos JK, Hillner BE, Sanyal AJ. A decision analysis study of the value of a liver biopsy in nonalcoholic steatohepatitis. Liver Int 2008;28:650-658.

25. Collins GS, Altman DG. Predicting the 10 year risk of cardiovascular disease in the United Kingdom: independent and external validation of an updated version of QRISK2. BMJ 2012;344:e4181.

26. Siddiqui MM, Rais-Bahrami S, Turkbey B, George AK, Rothwax J, Shakir N, Okoro C, et al. Comparison of MR/ultrasound fusion-guided biopsy with ultrasound-guided biopsy for the diagnosis of prostate cancer. JAMA 2015;313:390-397.

27. Van Calster B, Vickers AJ, Pencina MJ, Baker SG, Timmerman D, Steyerberg EW. Evaluation of markers and risk prediction models: overview of relationships between NRI and decision-analytic measures. Med Decis Making 2013;33:490-501.

28. Castera L, Vergniol J, Foucher J, Le Bail B, Chanteloup E, Haaser M, Darriet M, et al. Prospective comparison of transient elastography, Fibrotest, APRI, and liver biopsy for the assessment of fibrosis in chronic hepatitis C. Gastroenterology 2005;128:343-350.

29. Boursier J, de Ledinghen V, Leroy V, Anty R, Francque S, Salmon D, Lannes A, et al. A stepwise algorithm using an at-a-glance first-line test for the non-invasive diagnosis of advanced liver fibrosis and cirrhosis. J Hepatol 2017;66:1158-1165.

30. Newsome PN, Cramb R, Davison SM, Dillon JF, Foulerton M, Godfrey EM, Hall R, et al. Guidelines on the management of abnormal liver blood tests. Gut 2018;67:6-19.

31. Tsochatzis EA, Newsome PN. Non-alcoholic fatty liver disease and the interface between primary and secondary care. Lancet Gastroenterol Hepatol 2018;3:509-517.

32. D'Amico G, Garcia-Tsao G, Pagliaro L. Natural history and prognostic indicators of survival in cirrhosis: a systematic review of 118 studies. J Hepatol 2006;44:217-231.

This article is protected by copyright. All rights reserved.

Accepted Article

33. Zhang BH, Yang BH, Tang ZY. Randomized controlled trial of screening for hepatocellular carcinoma. J Cancer Res Clin Oncol 2004;130:417-422.

34. Bruix J, Sherman M, American Association for the Study of Liver D. Management of hepatocellular carcinoma: an update. Hepatology 2011;53:1020-1022.

35. Kovalak M, Lake J, Mattek N, Eisen G, Lieberman D, Zaman A. Endoscopic screening for varices in cirrhotic patients: data from a national endoscopic database. Gastrointest Endosc 2007;65:82-88.

36. Ioannou GN, Splan MF, Weiss NS, McDonald GB, Beretta L, Lee SP. Incidence and predictors of hepatocellular carcinoma in patients with cirrhosis. Clin Gastroenterol Hepatol 2007;5:938-945, 945 e931-934.

37. Poynard T, Cales P, Pasta L, Ideo G, Pascal JP, Pagliaro L, Lebrec D. Beta-adrenergic-antagonist drugs in the prevention of gastrointestinal bleeding in patients with cirrhosis and esophageal varices. An analysis of data and prognostic factors in 589 patients from four randomized clinical trials. Franco-Italian Multicenter Study Group. N Engl J Med 1991;324:1532-1538.

38. West J, Card TR. Reduced mortality rates following elective percutaneous liver biopsies. Gastroenterology 2010;139:1230-1237.

39. Silvis SE, Nebel O, Rogers G, Sugawa C, Mandelstam P. Endoscopic complications. Results of the 1974 American Society for Gastrointestinal Endoscopy Survey. JAMA 1976;235:928-930.

Accepted Article

This article is protected by copyright. All rights reserved.

Table 1: Decision tree input data

Input Variable	Study	Input Data	Follow-up interval	Level of Evidence	Comment
Probability of oesophageal varices when cirrhosis is present	Kovalak et al(35)	52.19% (881/1688)	Cross-sectional data only	Retrospective observational study	Interrogation of US national endoscopy database from 2000-2003
Probability of HCC when cirrhosis is present	Ioannou et al(36)	4.70% (100/2126)	3.6 years mean follow-up	Retrospective observational study	Interrogation of US Veteran's Affairs care database from 1994-2003
Probability of death if oesophageal varices are missed	Poynard et al(37)	18.36% (55/303)	2 years	Systematic review and meta-analysis of RCTs	Pooled data from 4 RCTs, 589 participants
Probability of death if oesophageal varices are diagnosed early	Poynard et al(37)	10.42% (29/286)	2 years	Systematic review and meta-analysis of RCTs	Pooled data from 4 RCTs, 589 participants
Probability of death from HCC in non-screened patients	Zhang et al(33)	79.71% (54/67)	No median follow-up available in non-screened group	Cluster RCT	HBV screening population of 18,816
Probability of death from HCC in screened patients	Zhang et al(33)	37.5% (32/86)	2 year median screening follow-up	Cluster RCT	HBV screening population of 18,816
Probability of death due to liver biopsy	West et al(38)	0.01% (11/61,187)	Cross-sectional data only	Retrospective observational study	UK NHS record data linkage from 1998-2005
Probabilty of death due to upper GI endoscopy	Silvis et al(39)	0.004% (10/211,410)	Cross-sectional data only	Retrospective survey	Survey of 404 ASGE members from 1972-1973

Abbreviations: HCC, hepatocellular carcinoma; RCT, randomised control trial; HBV, hepatitis B virus; NHS, National Health Service; ASGE, American Society of Gastrointestinal Endoscopy

This article is protected by copyright. All rights reserved.

Accepted Article

Table 2: Results of the decision tree model (N= 1000)

Prevalence	Test Strategy	Sens (%)	Spec (%)	TP	FP	FN	TN	CC (%)	Mortality (%)	Equal or lower mortality than biopsy
5%	Liver biopsy	95	100	47	0	3	950	95	0.41	Reference
	Decision tree mNIT	89	75	44	238	6	712	76	0.41	Reference
	Combined mNIT*	89	88	45	114	5	836	88	0.40	Reference
	Single Tests									
	APRI LS	75	78	38	209	12	741	78	0.45	No
	APRI HS	45	93	23	66	28	883	91	0.54	No
	FIB-4 LS	84	71	42	275	8	675	72	0.42	No
	FIB-4 HS	42	92	21	76	29	874	90	0.55	No
	FibroTest® LS	89	65	44	332	6	618	66	0.41	Yes
	FibroTest® HS	73	94	36	57	14	893	93	0.45	No
	VCTE	89	89	45	104	5	846	89	0.40	Yes
	Sequential Tests									
	APRI LS + VCTE	68	98	34	23	16	927	96	0.47	No
	APRI HS + VCTE	94	83	47	163	3	787	83	0.39	Yes
	APRI Dual + VCTE	73	91	36	82	14	868	90	0.45	No
	FIB-4 LS + VCTE	75	97	37	30	13	920	96	0.45	No
	FIB-4 HS + VCTE	94	82	47	172	3	778	92	0.39	Yes
	FIB-4 Dual + VCTE	80	90	40	98	10	852	89	0.43	No
	FibroTest® LS + VCTE	78	96	39	38	11	912	95	0.44	No
	FibroTest® HS + VCTE	98	84	49	84	1	795	84	0.38	Yes
	FibroTest® Dual + VCTE	87	91	43	87	7	863	91	0.41	Yes
20%	Liver biopsy	95	100	190	0	10	800	99	1.54	Reference
	Decision tree mNIT	94	63	187	296	13	504	69	1.54	Reference
	Combined mNIT	94	85	188	120	12	680	87	1.53	Reference
	Single Tests									
	APRI LS	75	78	150	176	50	624	77	1.76	No
	APRI HS	45	93	90	56	110	744	83	2.13	No
	FIB-4 LS	84	71	168	232	32	568	74	1.66	No
	FIB-4 HS	42	92	84	64	116	736	82	2.16	No
	FibroTest® LS	89	65	178	280	22	520	70	1.60	No
	FibroTest® HS	73	94	146	48	54	752	90	1.79	No
	VCTE	89	89	178	88	22	712	89	1.59	No
	Sequential Tests									
	APRI LS + VCTE	67	98	133	19	67	781	92	1.86	No
	APRI HS + VCTE	94	83	188	138	12	662	85	1.53	Yes
	APRI Dual + VCTE	72	91	143	69	57	731	87	1.80	No
	FIB-4 LS + VCTE	75	97	150	26	50	774	93	1.76	No
	FIB-4 HS + VCTE	94	82	187	145	13	655	84	1.53	Yes
	FIB-4 Dual + VCTE	79	90	159	82	41	718	88	1.71	No
	FibroTest® LS + VCTE	79	96	158	31	42	769	93	1.71	No
	FibroTest® HS + VCTE	97	84	194	131	6	669	86	1.50	Yes
	FibroTest® Dual + VCTE	87	91	174	72	26	728	90	1.62	No
	Liver biopsy	95	100	425	0	75	500	93	3.81	Reference
	mNIT^	94	87	472	65	28	435	91	3.81	Reference
	Single Tests									
	APRI LS	75	78	375	110	125	390	77	4.39	No

This article is protected by copyright. All rights reserved.

50%	APRI HS	45	93	225	35	275	465	69	5.30	No
	FIB-4 LS	84	71	420	145	80	355	84	4.12	No
	FIB-4 HS	42	92	210	40	290	460	67	5.39	No
	FibroTest® LS	89	65	445	175	55	325	77	3.97	No
	FibroTest® HS	73	94	365	30	135	470	84	4.45	No
	VCTE	89	89	445	55	55	445	89	3.97	No
	Sequential Tests									
	APRI LS + VCTE	67	98	334	12	166	488	83	4.63	No
	APRI HS + VCTE	94	83	470	86	30	414	88	3.82	No
	APRI Dual + VCTE	72	91	359	43	142	457	82	4.48	No
	FIB-4 LS + VCTE	75	97	374	16	126	484	86	4.39	No
	FIB-4 HS + VCTE	94	82	468	91	32	409	88	3.82	No
	FIB-4 Dual + VCTE	79	90	397	52	103	448	85	4.27	No
	FibroTest® LS + VCTE	79	96	396	19	104	481	88	4.27	No
	FibroTest® HS + VCTE	97	84	485	82	15	418	90	3.73	Yes
	FibroTest® Dual + VCTE	87	91	436	46	64	454	89	4.03	No

*Specificity of mNIT adjusted using decision curve analysis to have equal or greater net benefit to liver biopsy at a threshold probability of 0.20. The mNIT sensitivity is from the decision tree model.

^Adjustment of the mNIT specificity from the decision tree model was not necessary as the net benefit was greater than liver biopsy at threshold probability of 0.30.

Abbreviations: Sens, sensitivity; Spec, specificity; TP, true positive; FP, false positive; FN, false negative; TN, true negative; CC, correctly classified; mNIT, minimum sensitivity and specificity for equivalent mortality between liver biopsy and non-invasive testing; APRI, AST to Platelet Ratio Index; VCTE, Vibration Controlled Transient Elastography; LS, low specificity; HS, high specificity; Dual, denotes both high and low specificity test cut-offs

Accepted Article

This article is protected by copyright. All rights reserved.

Table 3: Test strategies with the greatest net benefit at different threshold probabilities independent of the results of the decision tree model

P_t	5% Prevalence				20% Prevalence				50% Prevalence			
	Single	Net B	Sequential	Net B	Single	Net B	Sequential	Net B	Single	Net B	Sequential	Net B
0.05	VCTE	0.038	FibroTest® HS +VCTE	0.041	VCTE	0.17	FibroTest® HS +VCTE	0.19	mNIT	0.47	FibroTest® HS +VCTE	0.48
0.10	VCTE	0.032	FibroTest® LS +VCTE	0.035	VCTE	0.17	FibroTest® HS +VCTE	0.18	mNIT	0.46	FibroTest® HS +VCTE	0.48
0.15	FibroTest® HS	0.026	FibroTest® LS +VCTE	0.032	VCTE	0.16	FibroTest® HS +VCTE	0.17	mNIT	0.46	FibroTest® HS +VCTE	0.47
0.20	FibroTest® HS	0.022	FIB-4 LS +VCTE	0.030	Liver Biopsy	0.16	FibroTest® HS +VCTE	0.16	mNIT	0.46	FibroTest® HS +VCTE	0.46
0.25	FibroTest® HS	0.017	FIB-4 LS +VCTE	0.027	Liver Biopsy	0.16	Liver Biopsy	0.16	mNIT	0.44	FibroTest® HS +VCTE	0.46
0.30	Liver Biopsy	0.015	FIB-4 LS +VCTE	0.025	Liver Biopsy	0.16	Liver Biopsy	0.16	mNIT	0.44	FibroTest® HS +VCTE	0.45
0.40	Liver Biopsy	0.015	APRI LS +VCTE	0.019	Liver Biopsy	0.16	Liver Biopsy	0.16	Liver Biopsy	0.44	Liver Biopsy	0.44
0.50	Liver Biopsy	0.015	Liver biopsy	0.015	Liver Biopsy	0.16	Liver Biopsy	0.16	Liver Biopsy	0.44	Liver Biopsy	0.44

Abbreviations: P_t, threshold probability; Net B, net benefit; APRI, AST to Platelet Ratio Index; VCTE, Vibration Controlled Transient Elastography; mNIT, minimum sensitivity and specificity for equivalent mortality between liver biopsy and non-invasive testing; LS, low specificity; HS, high specificity

Accepted Article

This article is protected by copyright. All rights reserved.

Table 4: Test strategies with agreement between decision tree and decision curve models at a threshold probability of 0.20

Prevalence	Testing Strategy	Sens (%)	Spec (%)	TP	FP	FN	TN	CC (%)	Decision Tree Mortality (%)	P_t above which LB strategy has greater net-benefit	Fit for practice
5%	Liver biopsy	95	100	47	0	3	950	95	0.41	-	Reference
	Combined mNIT*	89	88	45	114	5	836	88	0.40	0.20	Reference
	VCTE	89	89	45	104	5	846	89	0.40	0.21	Yes
	FibroTest® Dual + VCTE	87	91	43	87	7	863	91	0.41	0.24	Yes
20%	Liver biopsy	95	100	190	0	10	800	99	1.54	-	Reference
	Combined mNIT*	94	85	188	120	12	680	87	1.53	0.20	Reference
	FibroTest® HS + VCTE	97	84	194	131	6	669	86	1.50	0.21	Yes
50%	Liver Biopsy	95	100	425	0	75	500	93	3.81	-	Reference
	mNIT^	94	87	472	65	28	435	91	3.81	0.30	Reference
	FibroTest® HS + VCTE	97	84	485	82	15	418	90	3.73	0.33	Yes

*Specificity of combined mNIT adjusted using decision curve analysis to have equal or greater net benefit to liver biopsy at threshold probability of 0.20. The mNIT sensitivity is from the decision tree model.

^Adjustment of the mNIT specificity from the decision tree model was not necessary as the net benefit was greater than liver biopsy at threshold probability of 0.30.

Abbreviations: Sens, sensitivity; Spec, specificity; TP, true positive; FP, false positive; FN, false negative; TN, true negative; CC, correctly classified; P_t, threshold probability; VCTE, Vibration Controlled Transient Elastography; mNIT, minimum sensitivity and specificity for equivalent mortality between liver biopsy and non-invasive testing; LS, low specificity; HS, high specificity; Dual, denotes both high and low specificity test cut-offs

Accepted Article

This article is protected by copyright. All rights reserved.

Accepted Article

Fig. 1: Decision tree for diagnosing cirrhosis using liver biopsy or non-invasive testing strategies

Abbreviations: NIT, non-invasive test; HCC, hepatocellular carcinoma

Fig. 2: Summary of methods to compare a non-invasive test to liver biopsy for diagnosing cirrhosis using decision tree model and decision curve analysis

Abbreviations: NIT, non-invasive test; mNIT, minimum sensitivity and specificity for equivalent mortality between liver biopsy and non-invasive testing; Sens, sensitivity; Spec, specificity; TP, true positive; FP, false positive; FN, false negative; TN, true negative; %mort, individual risk of death over 2-years for each testing strategy outcome; P_t, threshold probability; HCC, hepatocellular carcinoma

Fig. 3: Sequential non-invasive test scenarios

Abbreviations: APRI, AST to Platelet Ratio Index; VCTE, Vibration Controlled Transient Elastography
A: Serum test dual cut-off sequential strategy
B: Serum test high specificity cut-off sequential strategy
C: Serum test low specificity cut-off sequential strategy

Fig. 4: Decision curves for test strategies demonstrating agreement between decision models

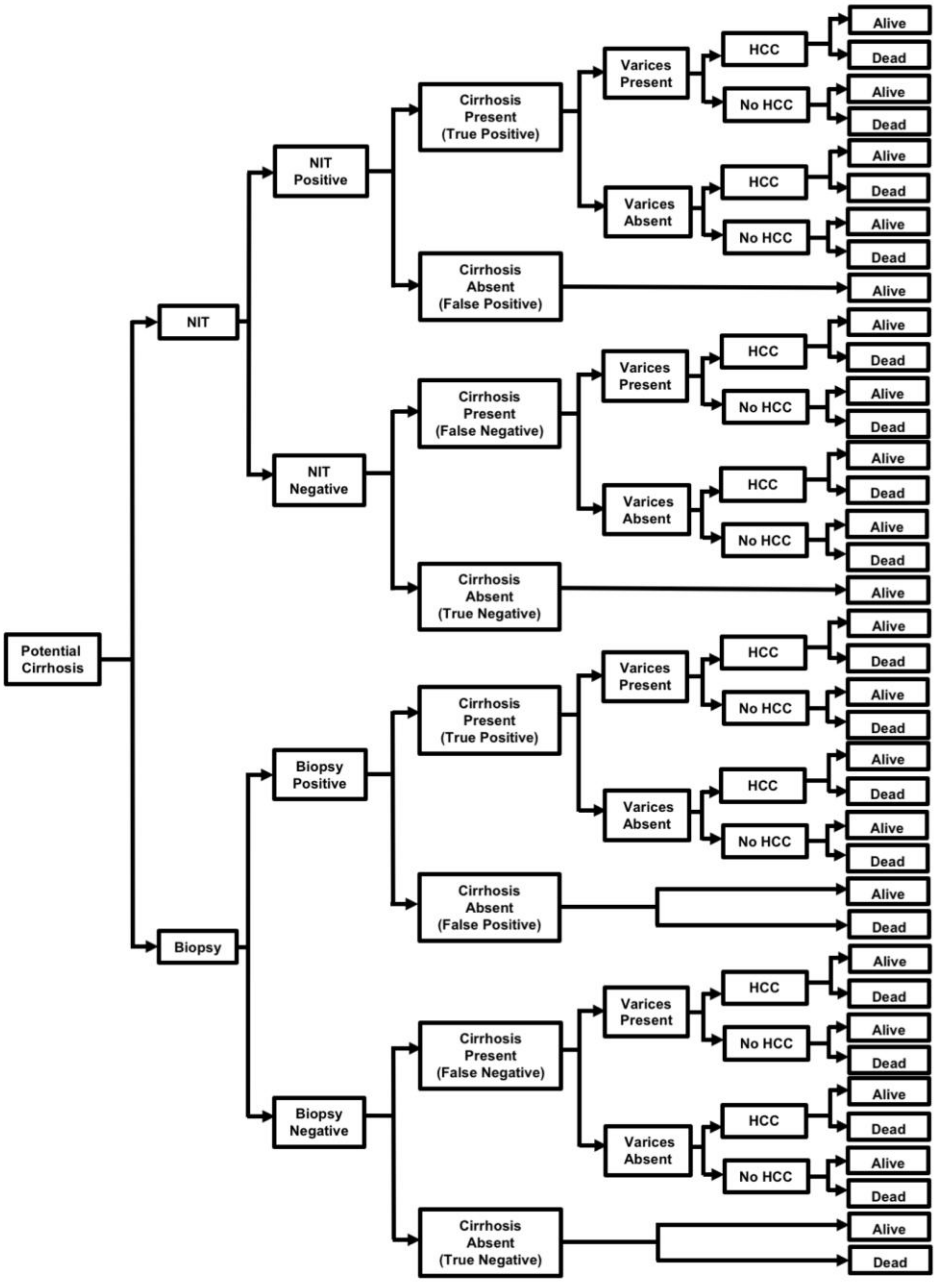
Decision curves representing test strategies across various threshold probabilities. These are presented in relation to treating all patients as having cirrhosis ("Treat All") and treating no patients as having cirrhosis ("Treat None"), which has no benefit. The intersection of the Treat All and Treat None curves corresponds to the prevalence of cirrhosis. The strategy that exhibits the greatest net benefit is the highest curve at each threshold probability. The liver biopsy sensitivity was 95% and specificity was 100%, resulting in a flat curve due to the absence of false positive diagnoses. A harm factor of 3.2% was applied to liver biopsy to account for invasiveness and reduced the net benefit proportionately. Decision Tree mNIT refers to the sensitivity and specificity of mNIT derived from the decision tree analysis. Combined mNIT represents the revised mNIT specificity derived from the decision curve analyses (Supplementary Figure 1) and optimised for a threshold probability of 0.20, with the same sensitivity as the decision tree model. NITs that had lower mortality than liver biopsy on the decision tree and higher net benefit on decision curve analysis are presented here.

A: 5% cirrhosis prevalence
B: 20% cirrhosis prevalence.
C: 50% cirrhosis prevalence. As the mNIT specificity derived from the decision tree had greater net benefit than biopsy at the probability threshold of 0.20, no additional adjustment to the mNIT specificity was required.

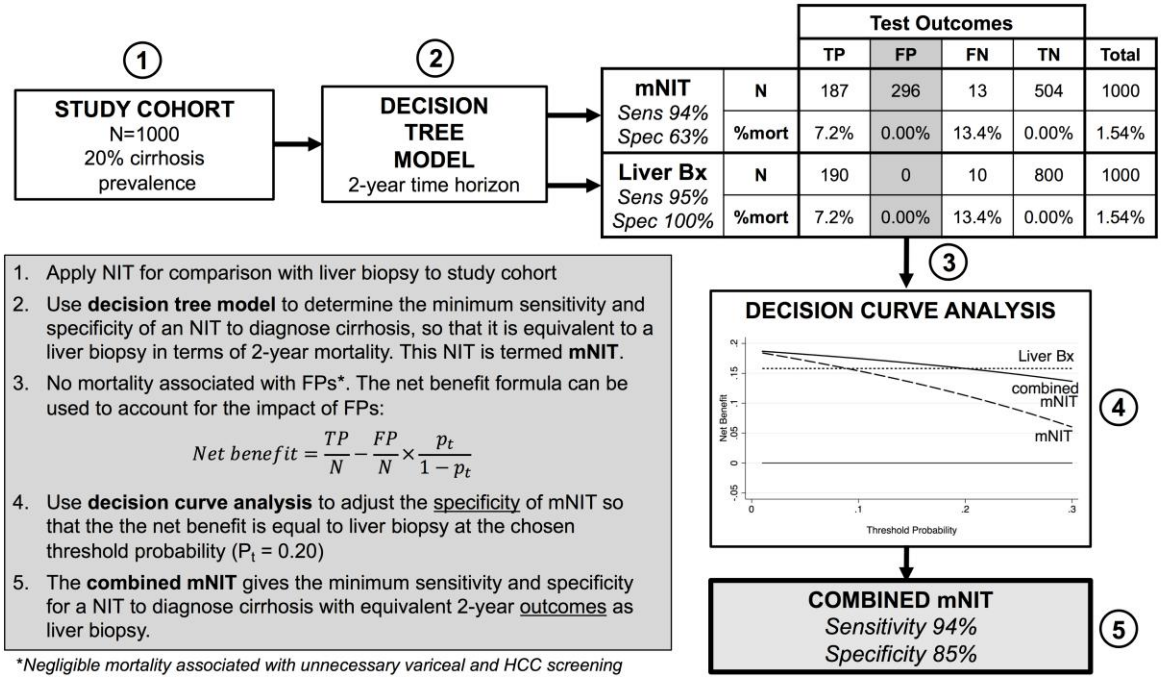
Abbreviations: VCTE, Vibration Controlled Transient Elastography; mNIT, minimum sensitivity and specificity for equivalent mortality between liver biopsy and non-invasive testing; NIT, non-invasive test; HS, high specificity; Dual, denotes both high and low specificity test cut-offs

This article is protected by copyright. All rights reserved.

Accepted Article



This article is protected by copyright. All rights reserved.



1. Apply NIT for comparison with liver biopsy to study cohort
2. Use **decision tree model** to determine the minimum sensitivity and specificity of an NIT to diagnose cirrhosis, so that it is equivalent to a liver biopsy in terms of 2-year mortality. This NIT is termed **mNIT**.
3. No mortality associated with FPs*. The net benefit formula can be used to account for the impact of FPs:

$$Net\ benefit = \frac{TP}{N} - \frac{FP}{N} \times \frac{p_t}{1 - p_t}$$

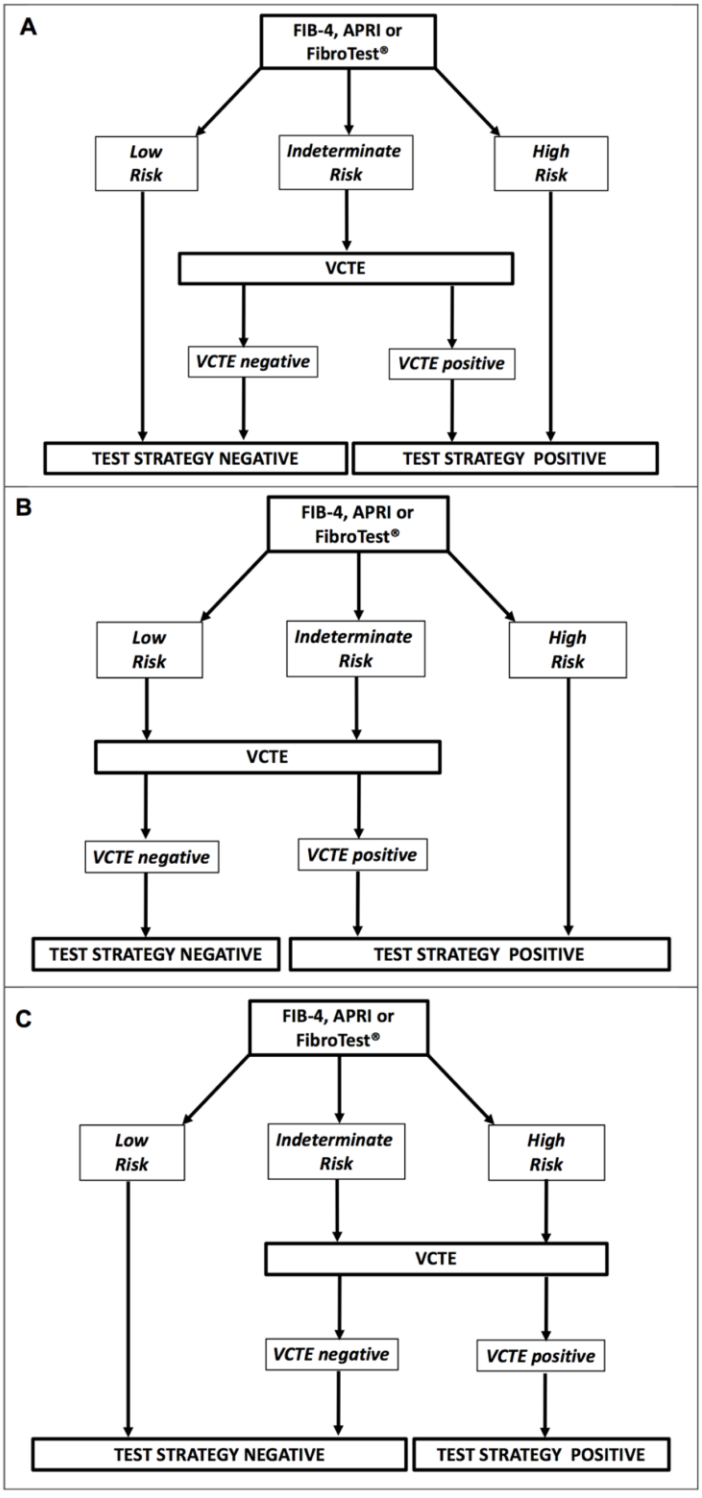
4. Use **decision curve analysis** to adjust the specificity of mNIT so that the the net benefit is equal to liver biopsy at the chosen threshold probability (P_t = 0.20)
5. The **combined mNIT** gives the minimum sensitivity and specificity for a NIT to diagnose cirrhosis with equivalent 2-year outcomes as liver biopsy.

**Negligible mortality associated with unnecessary variceal and HCC screening*

Accepted Article

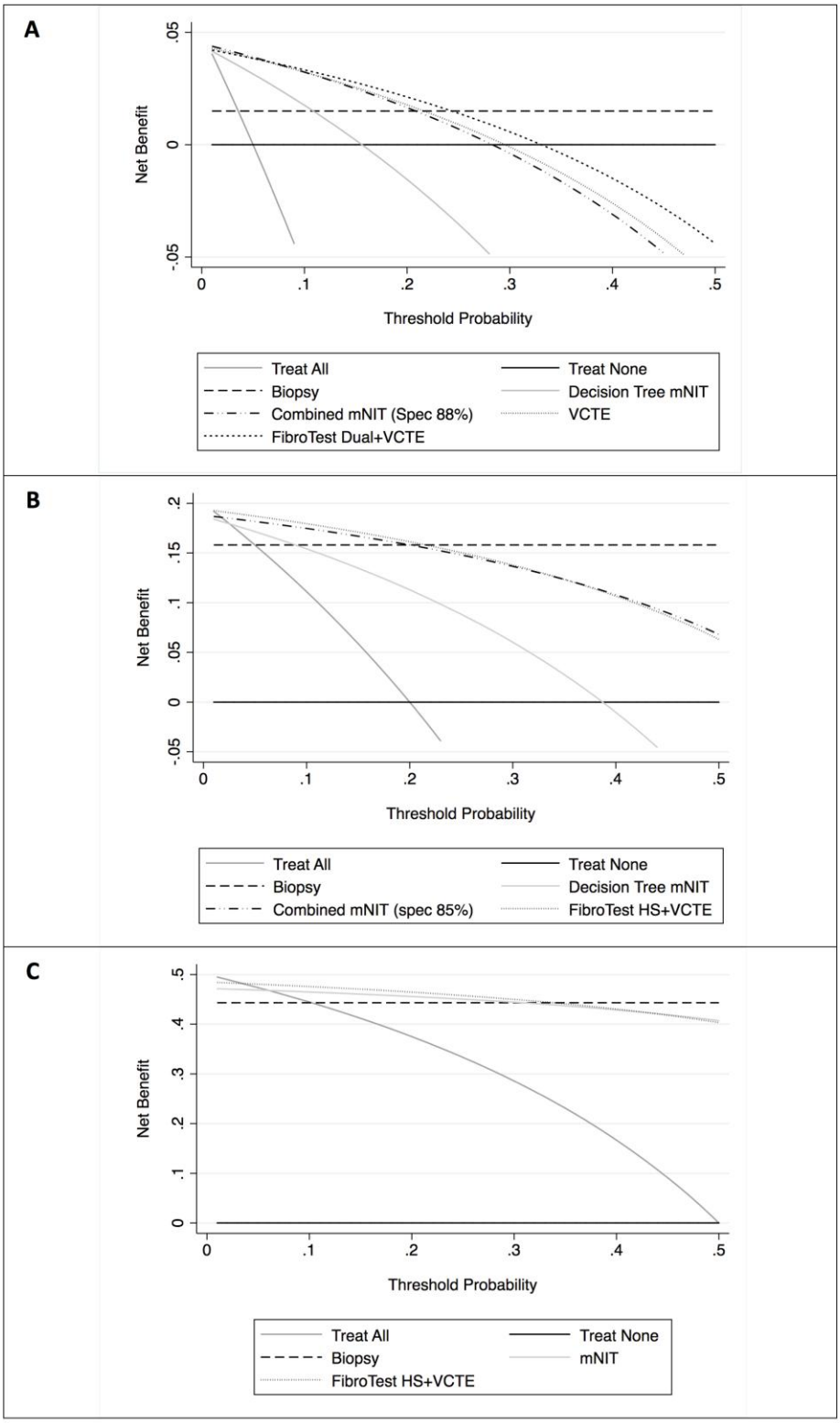
This article is protected by copyright. All rights reserved.

Accepted Article



This article is protected by copyright. All rights reserved.

Accepted Article



This article is protected by copyright. All rights reserved.