Technoleg Iechyd Cymru
Health Technology Wales

# Topic Exploration Report [1]

Topic explorations are designed to provide a high-level briefing on new topics submitted for consideration by Health Technology Wales. The main objectives of this report are to:

- Determine the quantity of evidence available for a technology of interest.
- Identify any gaps in the evidence.
- Inform decisions on topics that warrant fuller assessment by Health Technology Wales (HTW).

| | |
|---|---|
| **Topic exploration report number** | TER549 |
| **Topic** | Artificial intelligence-interpretation of laboratory results to predict fatty liver disease in primary care |
| **Summary of findings** | Fatty liver disease is a condition where excess fat builds up in the liver that can lead to fibrosis and scar tissue formation. It is asymptomatic in the initial stages and is reversible if caught early, however, it is often detected too late. Artificial intelligence (AI) can be used to predict the risk of liver disease using the results of routine clinical laboratory tests.<br><br>We identified two pieces of NICE guidance on the diagnosis of non-alcoholic fatty liver disease (NAFLD) and cirrhosis, one systematic review of AI techniques for NAFLD diagnosis, six validation studies of AI/machine learning approaches to diagnose fatty liver disease, one ongoing study, and one conference poster reporting economic evidence.<br><br>AI interpretation of routine laboratory tests and clinical data to diagnose fatty liver disease generally appears to have acceptable accuracy and provides similar or superior accuracy to currently used tests and fibrosis scoring indices. The only source of economic evidence suggests that the LiverPRO software is more accurate and less costly than the Enhanced Liver Fibrosis (ELF) test and Fibrosis-4 Index (FIB-4).<br><br>There is variability in the inputs used in the AI models and what aspects of liver disease they assess, with uncertainty around which models, other than LiverPRO, have regulatory approval. There is no evidence on clinical outcomes after use of these models and no economic evidence from a peer-reviewed source. |

[1] Cyfieithu dogfennau HTW wedi'u cyhoeddi o'r Saesneg i'r Gymraeg
Translation of published technical HTW documents from English into Welsh

## Introduction and aims

Steatotic liver disease, also known as fatty liver disease, is a condition where excess fat builds up in the liver. It is asymptomatic in the initial stages and is reversible if caught early, however, it is often detected too late. Subtypes of fatty liver disease include metabolic dysfunction-associated steatotic liver disease (MASLD), alcohol-associated liver disease (ALD), and metabolic and alcohol-associated liver disease (metALD). MASLD is often referred to as non-alcoholic fatty liver disease (NAFLD) as the terminology was changed recently in 2023.

There are several clinical parameters and serological tests that can act as indicators of potential fatty liver disease, even before symptoms are present. However, there is a lack of investigations suitable for use in primary care. Artificial intelligence (AI) can be used to assess the results of multiple routine laboratory tests and clinical data for a patient to indicate the risk of that person having fatty liver disease, potentially with increased accuracy compared to any of these factors alone.

LiverPRO is a CE-marked diagnostic platform for detecting fatty liver disease in primary care. The software's machine learning (ML) algorithms combine the results from at-risk peoples' standard blood tests to predict the level of scar tissue (fibrosis) in the liver. LiverPRO combines up to 10 standard biomarkers, depending on what data and test results are available, creating a liver fibrosis risk score shown as a percentage risk of having liver fibrosis. LiverPRO uses a set of multivariable models and selects the algorithm with the best accuracy based on the data available for that patient (Lindvig et al. 2024). The biomarkers included in the model include nine biochemical markers (aspartate aminotransferase, gamma-glutamyl transferase, alkaline phosphatase, total cholesterol, sodium, international normalised ratio [INR], bilirubin, albumin, platelets), and age. LiverPRO also provides a personalised clinical recommendation to aid GPs with decision-making and management of the patient.

Health Technology Wales researchers searched for evidence on the clinical effectiveness and cost effectiveness of AI-interpretation of laboratory blood sample results to predict liver fibrosis in primary care.

## Evidence overview

A similar topic, concerning diagnosis of liver disease, was the subject of a previous HTW topic exploration report: TER506 ELF test (Enhanced Liver Fibrosis) (HTW 2024). The ELF test is currently undergoing full appraisal by HTW.

AI-interpretation of laboratory blood sample results to predict liver fibrosis is a digital health technology and was determined to be a Tier C technology according to the Evidence Standards Framework for Digital Health Technologies. Technologies within this classification provide information that will be used to take an immediate or near-term action to diagnose, screen or detect a disease or condition.

To demonstrate effectiveness of the technologies of this classification, it is recommended that one or more high-quality studies are done using the technology to support the claimed benefits of the test. This may include test accuracy studies, using an appropriate reference standard, or a concordance study to show agreement with current practice. Relevant outcomes may include test accuracy and time to diagnosis (if this is a claimed benefit of the test). Test accuracy alone does not demonstrate clinical utility and may need to be linked to existing studies (including studies on other technologies) reporting the downstream clinical consequences of the diagnosis or test outcome.

## Evidence overview

### Guidance

HTW researchers identified two relevant pieces of guidance from NICE regarding liver disease. NICE guideline NG49, covering the assessment and management of non-alcoholic fatty liver disease (NICE 2016a), recommends not to use routine liver blood tests to rule out NAFLD in higher risk groups or to assess for advanced liver fibrosis in people with NAFLD. It recommends that the ELF test should be considered to test for advanced liver fibrosis in people with NAFLD. NG50, covering the assessment and management of cirrhosis in people aged 16 years or older (NICE 2016b), also recommends not to use routine liver blood tests to rule out cirrhosis. Neither of these guidelines mentions using AI to aid diagnosis or prediction of liver damage, likely due to the recency of AI tools development.

### Secondary evidence

One systematic review of AI techniques for NAFLD diagnosis was identified. Zamanian et al. (2024) included 28 studies that used AI algorithms to diagnose diseases related to NAFLD by using clinical data and serological tests. Features included in the models included age, sex, race, diabetes status, body mass index (BMI), cholesterol, blood cell counts, alanine transferase, aspartate transferase, and bilirubin, amongst others. The AI models used differentiated between NAFLD vs. no NAFLD, non-alcoholic steatohepatitis (NASH) vs. no NASH, NAFLD vs. NASH, stage of NASH, degree of steatosis, or degree of fibrosis. The accuracy of the various models was judged to be acceptable by the authors, with areas under the receiver-operator curve (AUC) ranging from 0.64 to 0.95 and accuracy ranging from 0.52 to 0.9989.

### Primary evidence

HTW researchers identified five primary studies published after the search dates of the systematic review discussed above and also included one study that was provided by the topic proposer. All of the identified studies are validation studies.

Hassoun et al. (2023) developed a ML algorithm to identify severe liver fibrosis using a US public database, with a control population of 6,828 and a target population of 437 (Fibroscan values greater than or equal to 9.7 KPa). Liver biopsy was used as the reference standard. The authors identified the 26 most significant parameters to use in their model and validation testing of 300 subjects (150 controls and 150 target population) found that three oversampling methods lead to higher sensitivity (range 0.54 ± 0.03 to 0.68 ± 0.04), higher specificity (range 0.83 ± 0.01 to 0.88 ± 0.02), and higher AUC (all 0.87 ± 0.01) than undersampling methods. Two of the undersampling methods resulted in zero sensitivity. The authors conclude that the ML approach maintained high specificity and improved sensitivity compared to currently used clinical scoring systems.

Nabrdalik et al. (2024) developed a ML approach to assess the risk of MASLD in people with diabetes. Eight parameters were included in the model: age, BMI, type of diabetes, alanine aminotransferase, aspartate aminotransferase, platelet count, hyperuricaemia, and treatment with metformin. The model was trained on a sample of 1,735 participants and then validated in a set of 265 patients. The model achieved a sensitivity of 0.80, specificity of 0.74, and an AUC of 0.81 (95% confidence interval [CI] 0.76 to 0.87).

Another study developed multiple ML algorithms from a European patient dataset to predict MASLD outcomes such as metabolic dysfunction associated steatohepatitis, fibrosis, or cirrhosis using routine non-invasive clinical tests (McTeer et al. 2024). In the training set, AUCs were generated ranging from 0.719 to 0.994 for various MASLD outcomes. It was found that increasing the number of variables included in the models did not significantly improve predictive performance.

## Evidence overview

In a study by Verma et al. (2024), 21 ML models were developed in a cohort of Asian people with MASLD to detect significant liver fibrosis (greater than or equal to F2 fibrosis). The authors state the developed ML models were 7 to 12% better at discriminating significant fibrosis than Fibrosis-4 Index (FIB-4), with optimised random forest (RF) producing a negative predictive value and F1 score of 0.852 and 0.559, respectively, in the validation set. RF was determined to detect 10 times more patients with significant fibrosis, reduce unnecessary referrals by 28%, and reduce missed referrals by 78% compared to FIB-4. Features included in the RF model included age, diabetes status, BMI, cholesterol, HbA1c, alanine transferase, aspartate transferase, gamma glutamyl transferase, and bilirubin, amongst others.

A pre-print study by Lindvig et al. (2024), specifically examining the use of LiverPRO to diagnose steatotic liver disease, was also identified. The ML algorithms were developed in a cohort of 462 Danish participants, including those with liver disease related to metabolic dysfunction and alcohol. The models were validated in four independent cohorts (n = 8,586). The reference standard was transient elastography (TE) and LiverPRO detected significant fibrosis (TE greater than or equal to 8 KPa) with good accuracy (AUC 0.81, 95% CI 0.79 to 0.84) in one of the validation cohorts, similar to the ELF test and better than FIB-4 and NAFLD Fibrosis Score. Accuracy was slightly lower (AUC range 0.69 to 0.72) in the other validation cohorts. LiverPRO was found to correctly classify 82% of participants with significant fibrosis when a rule-out cut-off of less than 25% risk score was used and 95% of participants when a rule-in cut-off of greater than 65% risk score was used. Prognostic performance for liver related events was found to be strong when evaluated in a cohort of 470,795 participants from the UK Biobank.

Another validation study published before the search date of the systematic review by Zamanian et al. (2024), but not included in the review, was provided by the topic proposer. Six ensemble learning models of increasing complexity (LiverAID, the precursor to LiverPRO) were developed by Blanes-Vidal et al. (2022) from a cohort of 3,352 participants. These six models had AUCs ranging from 0.86 to 0.94 for identifying liver stiffness greater than 8 KPa and outperformed blood-based indices such as FIB-4.

### Economic evidence

A cost-comparison analysis reported in a conference poster was provided by the topic proposer. No other economic evidence was identified in HTW's searches. The poster by Lindvig et al. (2023) reported a comparison of using LiverPRO in primary care to FIB-4 and the ELF test in a prospective cohort study of 6,032 participants using a hospital perspective. The authors report that the patient pathway using LiverPRO as the first-tier test correctly classified 93% of participants compared with 85% and 63% for the ELF test and FIB-4 pathways, respectively. The mean costs per patient for the three pathways were reported as €26, €103, and €97 for LiverPRO, ELF test and FIB-4, respectively. These results indicated that LiverPRO was more accurate and less costly than the other two tests.

### Ongoing studies

One relevant ongoing study was identified. NCT06061640 is a case-control study being carried out in China to develop a ML prediction model to identify MASLD from metabolic indicators. The study is due to complete in December 2024.

## Areas of uncertainty

- There is variability in the published models, including what inputs they use to generate a diagnosis and what conditions or aspects of liver disease they assess.

## Areas of uncertainty

- It is uncertain whether any of the AI models in the included studies, besides LiverPRO, have regulatory approval.
- Various reference standards were used in the included studies, including Fibroscan, ultrasonography, and liver biopsy.
- There is no evidence on clinical outcomes after the use of AI models for diagnosing liver disease.
- All of the available evidence comes from validation studies.
- No peer reviewed economic evidence was identified.

## Literature search results

| Health technology assessments and guidance |
| --- |
| HTW. (2024). ELF test (Enhanced Liver Fibrosis). Topic exploration report TER506. Health Technology Wales. Available at: https://healthtechnology.wales/reports-guidance/elf-test-enhanced-liver-fibrosis/ [Accessed 07 May 2024].<br><br>NICE. (2016a). Non-alcoholic fatty liver disease (NAFLD): assessment and management. NICE guideline NG49. National Institute for Health and Care Excellence. Available at: https://www.nice.org.uk/guidance/ng49 [Accessed 07 May 2024].<br><br>NICE. (2016b). Cirrhosis in over 16s: assessment and management. NICE guideline NG50. National Institute for Health and Care Excellence. Available at: https://www.nice.org.uk/guidance/ng50 [Accessed 09 May 2024]. |
| **Evidence reviews and economic evaluations** |
| Zamanian H, Shalbaf A, Zali MR, et al. (2024). Application of artificial intelligence techniques for non-alcoholic fatty liver disease diagnosis: A systematic review (2005-2023). Comput Methods Programs Biomed. 244: 107932. doi: https://doi.org/10.1016/j.cmpb.2023.107932 |
| **Individual studies (published after search date of Zamanian et al. (2024)** |
| Hassoun S, Bruckmann C, Ciardullo S, et al. (2023). Setting up of a machine learning algorithm for the identification of severe liver fibrosis profile in the general US population cohort. International Journal of Medical Informatics. 170: 104932. doi: https://doi.org/10.1016/j.ijmedinf.2022.104932<br><br>Lindvig KP, Thorhauge KH, Hansen JK, et al. (2024). LiverPRO for the Prediction of Significant Liver Fibrosis in Primary Care: Development, Validation, and Prognostic Evaluation of a Novel Score. Preprints with The Lancet. doi: https://doi.org/10.2139/ssrn.4745064<br><br>McTeer M, Applegate D, Mesenbrink P, et al. (2024). Machine learning approaches to enhance diagnosis and staging of patients with MASLD using routinely available clinical information. PLoS One. 19(2): e0299487. doi: https://doi.org/10.1371/journal.pone.0299487<br><br>Nabrdalik K, Kwiendacz H, Irlik K, et al. (2024). Machine learning identifies metabolic dysfunction associated steatotic liver disease in patients with diabetes mellitus. J Clin Endocrinol Metab. doi: https://doi.org/10.1210/clinem/dgae060<br><br>Verma N, Duseja A, Mehta M, et al. (2024). Machine learning improves the prediction of significant fibrosis in Asian patients with metabolic dysfunction-associated steatotic liver disease - The Gut and Obesity in Asia (GO-ASIA) Study. Aliment Pharmacol Ther. 59(6): 774-88. doi: https://doi.org/10.1111/apt.17891 |
| **Ongoing research** |
| NCT06061640. The Potential Value and Impact of Diagnostic Biomarkers for MAFLD Using Machine Learning Methods. Available at: https://clinicaltrials.gov/study/NCT06061640?tab=table [Accessed 13 May 2024]. |
| **Provided by Topic Proposer** |
| Blanes-Vidal V, Lindvig KP, Thiele M, et al. (2022). Artificial intelligence outperforms standard blood-based scores in identifying liver fibrosis patients in primary care. Scientific Reports. 12(1): 2914. doi: https://doi.org/10.1038/s41598-022-06998-8<br><br>Lindvig KP, Kjærgaard M, Thorhauge K, et al. (2023). A cost-comparison of the LiverPRO score with FIB-4, ELF, and FibroScan in 6,032 study participants. Journal of Hepatology. 78: S163-S4. doi: https://doi.org/10.1016/S0168-8278(23)00679-7 |

| | |
|---|---|
| **Date of search** | 07 May 2024 |
| **Concepts used** | Liver disease, liver fibrosis, steatotic liver disease, SLD, fatty liver disease, hepatology, AI, artificial intelligence, machine learning, LiverPRO |

# Proposed research question and evidence selection criteria (if selected)

| | |
|---|---|
| **Proposed Research question** | **What is the clinical effectiveness and cost effectiveness of AI-interpretation of laboratory blood sample results to predict liver fibrosis in primary care?** |

| | Inclusion criteria | Exclusion criteria |
|---|---|---|
| **Population** | People at risk of fatty liver disease | |
| **Intervention** | AI-interpretation of routine laboratory results and clinical data in primary care | AI models that include test results and assessments that would need to take place in secondary care |
| **Comparison/ Comparators** | Standard of care: hepatic function panel (also known as liver blood tests, liver function tests, liver panel or hepatic function tests) alone<br>FIB-4<br><br>For diagnostic accuracy reference standards: liver biopsy, ultrasonography, transient elastography | |
| **Outcome measures** | Diagnostic accuracy outcomes (e.g., sensitivity, specificity, positive predictive value, negative predictive value)<br>Time to diagnosis<br>Time to treatment<br>Mortality<br>Changes in referrals to secondary care<br>Health related QoL<br>Resource use<br>Economic outcomes | |

| | |
|---|---|
| **Proposed speciality** | **Endocrine, nutritional and metabolic** |