

Protocol

# An Explainable AI Tool (FibroX) for Detecting Advanced Liver Fibrosis in Adults With Type 2 Diabetes: Protocol for a Pilot Crossover Trial

Basile Njei<sup>1,2,3,4,5,6</sup>, MD, MPH, PhD; Ulrick Sidney Kanmounye<sup>7</sup>, MD, MPH, MSc

<sup>1</sup>Section of Digestive Diseases, Department of Medicine, Yale University, New Haven, CT, United States

<sup>2</sup>Engelhardt School of Global Health and Bioethics, Euclid University, Bangui, Central African Republic

<sup>3</sup>VA Connecticut Healthcare, West Haven, CT, United States

<sup>4</sup>Ohio University Heritage College of Osteopathic Medicine, Athens, OH, United States

<sup>5</sup>Yale Liver Center, Yale New Haven Health, New Haven, CT, United States

<sup>6</sup>International Medicine Program, Yale University, New Haven, CT, United States

<sup>7</sup>Association of Future African Neurosurgeons, Yaounde, Cameroon

**Corresponding Author:**

Basile Njei, MD, MPH, PhD  
Section of Digestive Diseases  
Department of Medicine  
Yale University  
333 Cedar Street  
New Haven, CT, 06510  
United States  
Phone: 1 475 227 553  
Email: [basilenjei@gmail.com](mailto:basilenjei@gmail.com)

## Abstract

**Background:** Metabolic dysfunction–associated steatotic liver disease is highly prevalent in adults with type 2 diabetes, and advanced fibrosis is its strongest prognostic marker. However, existing noninvasive tools may underperform in diabetes care and are inconsistently used in practice.

**Objective:** This study aims to evaluate the feasibility, usability, and preliminary diagnostic effectiveness of FibroX, an explainable artificial intelligence tool for identifying metabolic dysfunction–associated steatotic liver disease–associated advanced fibrosis in adults with type 2 diabetes.

**Methods:** This 12-month health care provider–level randomized crossover pilot trial will enroll at least 36 primary care clinicians managing adults with type 2 diabetes. Participants will complete 2 simulated care periods: one with FibroX-enabled care and one with usual care, separated by a 1-week washout. FibroX generates individualized advanced fibrosis risk estimates from routine clinical data, provides guideline-aligned triage categories, and displays case-level explanatory outputs using Shapley Additive Explanations. FibroX is an investigational tool and is not currently used in routine clinical practice. Primary outcomes are feasibility and usability, while secondary and exploratory outcomes include workflow efficiency, preliminary diagnostic performance, and implementation measures.

**Results:** Institutional review board approval has been obtained. The protocol was registered on ClinicalTrials.gov (NCT07305324) on December 1, 2025. At the time of submission, recruitment had not yet begun. The study is currently unfunded. Preparatory activities, including platform development and case validation, are ongoing. Recruitment is expected to begin in June 2026, with primary completion anticipated in May 2027 and results expected to be available in late 2027.

**Conclusions:** This pilot study will provide preliminary evidence on the feasibility, usability, and diagnostic performance of an explainable artificial intelligence tool for fibrosis risk stratification in diabetes care and will inform the design of a future larger trial.

**Trial Registration:** ClinicalTrials.gov NCT07305324; <https://clinicaltrials.gov/study/NCT07305324>

**International Registered Report Identifier (IRRID):** PRR1-10.2196/90456

**KEYWORDS**

artificial intelligence; AI; liver fibrosis; metabolic dysfunction–associated steatotic liver disease; MASLD; pilot; randomized controlled trial; type 2 diabetes

## Introduction

### Background

Metabolic dysfunction–associated steatotic liver disease (MASLD) is highly prevalent and clinically consequential among people living with type 2 diabetes (T2D). Population studies indicate that steatosis and progressive fibrosis affect a large proportion of adults with T2D, with MASLD prevalence approaching 70% in some cohorts, and advanced fibrosis (stage  $\geq$ F3) conferring markedly elevated risks of liver-related events and cardiovascular morbidity and mortality [1-4]. Advanced fibrosis is the strongest prognostic marker in MASLD and is associated with 2 to 3-fold higher liver-related and cardiovascular mortality in diabetes populations [4,5]. Recent diabetes-focused guidance has therefore emphasized that adults with prediabetes or type 2 diabetes are a priority population for structured noninvasive fibrosis risk stratification rather than opportunistic case finding alone [2,3]. However, real-world implementation in primary care and diabetes clinics remains inconsistent, leading to missed opportunities for early intervention [1-4].

Current approaches to fibrosis assessment rely on a mix of laboratory-based composite scores (eg, the Fibrosis-4 Index [FIB-4] and nonalcoholic fatty liver disease fibrosis score) and imaging modalities (eg, vibration-controlled transient elastography [VCTE] and magnetic resonance elastography) [6-8]. While elastography provides strong diagnostic performance, it requires dedicated equipment and trained personnel, limiting availability; meanwhile, fixed-formula laboratory scores can underperform in T2D populations where metabolic and inflammatory profiles differ, and they do not capture complex, nonlinear relationships among routinely collected clinical variables [9-12]. Liver biopsy, although a reference standard, is invasive, subject to sampling variability, and not suitable for screening at scale [6]. Consequently, many individuals with T2D and advanced fibrosis remain undetected until late disease stages, when therapeutic options are limited and outcomes are poorer [7,13]. In real-world practice, these pathways often depend on simplified cutoffs followed by second-line elastography, which can leave many at-risk patients in indeterminate categories and create access and workflow barriers in primary care [4,14,15].

Explainable artificial intelligence (AI) offers a path to more accurate, transparent, and workflow-compatible decision support in diabetes care. Prior work has shown that machine learning models leveraging routinely available data can improve discrimination for high-risk MASLD and fibrosis outcomes,

while post hoc explanation methods enhance interpretability and clinician trust [16-19]. Recent studies in liver disease and other clinical prediction settings have increasingly used explainable machine learning approaches to move beyond black-box risk estimates and provide patient-level feature attribution that may support clinician interpretation and adoption [20-22].

Building on this foundation, FibroX was developed to estimate individualized risk of advanced fibrosis ( $\geq$ F3) using common clinical inputs (age, aspartate aminotransferase [AST], alanine aminotransferase [ALT], platelets, hemoglobin A<sub>1c</sub> [HbA<sub>1c</sub>], BMI, creatinine or estimated glomerular filtration rate), with dual-threshold triage bands aligned to guidelines (rule out, indeterminate to elastography, and rule in to hepatology referral) and Shapley Additive Explanations (SHAP) to surface the features most responsible for each prediction [17,18]. In prior work, FibroX was derived using an explainable Extreme Gradient Boosting framework applied to population-based National Health and Nutrition Examination Survey data and subsequently evaluated in external biopsy-confirmed MASLD cohorts. Across these earlier evaluations, FibroX showed better discrimination for advanced fibrosis than conventional noninvasive scores such as FIB-4 and was also associated with prognostic stratification for longer-term outcomes. However, FibroX remains investigational, and its feasibility, usability, and influence on clinician decision-making have not yet been prospectively assessed in a clinical workflow context [15-18].

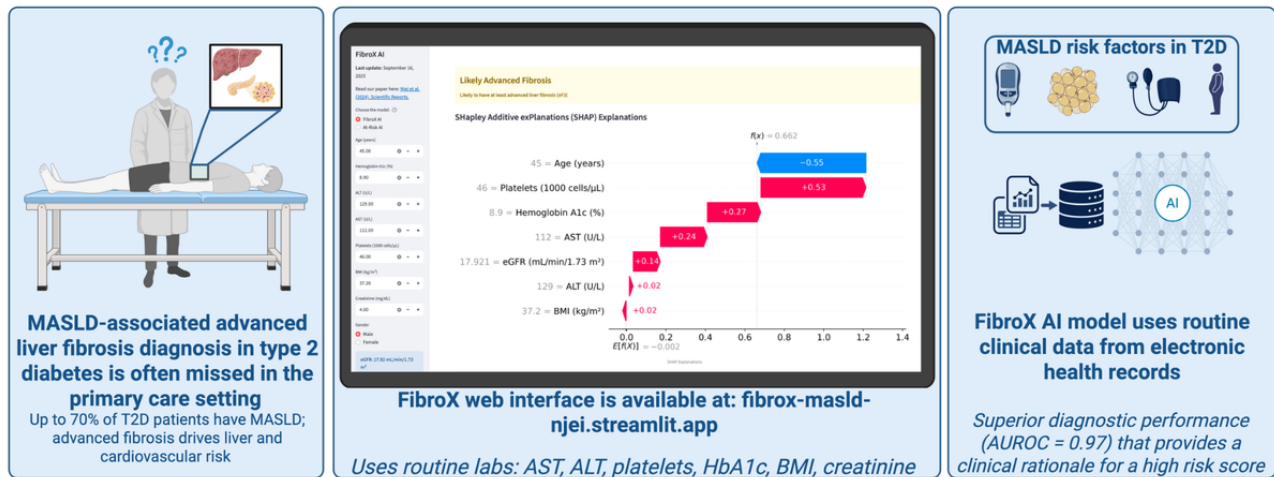
### Study Objectives and Rationale

This pilot aims to evaluate the feasibility, usability, and preliminary effectiveness of FibroX as an explainable AI decision support tool for detecting MASLD-associated advanced fibrosis in adults with T2D during simulated clinical encounters. Specifically, we aim to (1) assess the primary pilot end points of feasibility and usability in diabetes-focused primary care workflows using validated instruments, (2) estimate secondary outcomes related to workflow efficiency and preliminary diagnostic performance for advanced fibrosis, and (3) explore adoption and implementation metrics using the Reach, Effectiveness, Adoption, Implementation, and Maintenance framework, including guideline-concordant triage and net reclassification improvement (Figure 1).

If FibroX proves feasible and acceptable and improves diagnostic discrimination, it could help operationalize guideline-concordant triage, reduce missed diagnoses, and inform equitable MASLD screening strategies in diabetes care [1-4,10,12,13,16].

**Figure 1.** Schematic of FibroX inputs, explainable outputs, and clinician actions in the simulated diabetes care workflow. ALT: alanine aminotransferase; AST: aspartate aminotransferase; AUROC: area under the receiver operating characteristic curve; eGFR: estimated glomerular filtration rate; HbA<sub>1c</sub>: hemoglobin A1c; MASLD: metabolic dysfunction–associated steatotic liver disease; SHAP: Shapley Additive Explanations; T2D: type 2 diabetes.

## Accuracy and Acceptability of an Explainable AI Tool for Detecting MASLD-Associated Advanced Liver Fibrosis in Patients with Type 2 Diabetes: Provider-Level Crossover RCT Pilot



**Aim 1: Feasibility & Usability**  
*Q: Can and will diabetes care providers use FibroX?*

**Aim 2: Diagnostic Accuracy of Detecting Advanced Fibrosis in T2D patients with MASLD**  
*Q: Does FibroX improve diagnostic accuracy?*

**Aim 3: Implementation**  
*Q: How well does FibroX integrate into diabetes workflows?*

### Methods

#### Trial Design and Oversight

We will conduct a 12-month, health care provider–level, randomized, 2-period crossover simulation trial to evaluate the feasibility, usability, and preliminary effectiveness of FibroX, an explainable AI decision support tool for detecting MASLD-associated advanced liver fibrosis ( $\geq F3$ ) in adults with T2D. Because FibroX is not currently used in routine clinical care, this study evaluates the tool in a simulated clinical environment. Each clinician serves as their own control (FibroX-enabled care vs usual care), separated by a 1-week washout. Distinct, nonoverlapping simulation case sets will be used in each study period, such that clinicians will not evaluate the same cases across periods, reducing case-recall bias and minimizing carryover related to prior exposure. The study involves no patient enrollment and uses deidentified or synthetic cases. Oversight is by the principal investigator and study steering team; a formal Data Monitoring Committee is not required given the minimal risk.

#### Study Setting

Health care provider participants will be recruited from Yale-affiliated diabetes and primary care clinics and participating federally qualified health centers. Simulations are delivered via a secure, web-based platform designed to replicate an electronic health record (EHR)–like interface. Each simulated case presents standardized clinical information, including patient demographics, laboratory values, and relevant clinical notes. Participants review cases sequentially and make diagnostic and management decisions within this interface. In the FibroX-enabled condition, additional outputs—including predicted risk of advanced fibrosis, guideline-aligned triage categories, and SHAP-based feature explanations—are displayed alongside standard clinical data.

#### Participants

This pilot will enroll licensed primary care clinicians, including physicians, nurse practitioners, and physician assistants. Inclusion criteria require health care providers to have  $\geq 0.5$  full-time equivalent dedicated to adult diabetes care at participating sites, the ability to complete 2 simulation periods,

and provision of informed consent. Exclusion criteria include prior participation in any FibroX usability or validation studies, exclusive practice in pediatrics or nondiabetes specialty areas, or inability to complete all study procedures. Participating clinicians are expected to have varying levels of clinical experience, reflecting real-world primary care practice, but all are required to have active involvement in the management of adult patients with T2D to ensure familiarity with relevant clinical workflows and decision-making. Participants will be recruited through email invitations and departmental announcements across Yale-affiliated primary care and diabetes clinics, as well as participating federally qualified health centers. Recruitment will target clinicians with active involvement in the management of adults with T2D, with participation on a voluntary basis.

The simulation case library will comprise adult patients (aged  $\geq 18$  years) with T2D and at least one additional MASLD risk factor (eg, obesity, dyslipidemia, and hypertension). Each case must have a ground-truth fibrosis stage determined by liver biopsy or expert consensus based on VCTE. To ensure fidelity to real-world clinical presentation, cases will include complete routine clinical data such as age, AST, ALT, platelet count, HbA<sub>1c</sub>, BMI, and creatinine or estimated glomerular filtration rate, along with relevant clinical notes. The simulation case library will be developed using deidentified or synthetic cases derived from representative clinical datasets and prior cohorts, with case selection designed to capture a range of fibrosis risk profiles and clinical presentations. Cases will be curated to ensure clinical realism, balanced representation across fibrosis stages, and inclusion of all variables required for FibroX and comparator assessments.

### Interventions

The intervention arm, known as FibroX-enabled care, will use a simulated EHR-like interface delivered through a secure web-based platform that mirrors the presentation of routine outpatient clinical information. Participants will review standardized simulated cases containing patient demographics, relevant clinical history, laboratory data, and clinical notes. Within this interface, FibroX-enabled care will display key predictive and explanatory information at the point of decision-making. This will include individualized FibroX risk estimates for advanced fibrosis generated using routinely available clinical data. Crucially, the interface presents dual-threshold triage bands aligned with established clinical guidance, classifying patients into 3 actionable categories: rule out, indeterminate (with a pathway to VCTE or other noninvasive testing), and rule in (with a pathway to hepatology referral). In addition, the interface presents case-level SHAP outputs that display the relative contribution of individual clinical variables (eg, age, AST, ALT, platelet count, HbA<sub>1c</sub>, BMI, and kidney function) to the predicted risk of advanced fibrosis. These explanations are presented in a structured visual format, such as ranked feature contributions, alongside the risk estimate and triage category to support clinician interpretation at the point of decision-making. By providing case-level explanations of the key drivers of predicted risk, these SHAP-based outputs are intended to support clinician

interpretation of the model's recommendations, increase confidence in triage decisions, and facilitate more consistent application of guideline-concordant referral pathways compared with usual care, where risk estimation is often implicit and variable.

The comparator arm, designated as usual care, will present the same standardized clinical case information within the simulated EHR-like interface but without FibroX risk estimates, SHAP explanations, or AI-based triage support. Participants in this arm will make diagnostic and management decisions based on routine clinical judgment alone. Optional access to a basic FIB-4 calculator will be provided, reflecting common clinical practice for noninvasive fibrosis risk stratification. FIB-4 is a widely used blood-based fibrosis score derived from age, AST, ALT, and platelet count to estimate the likelihood of advanced fibrosis [23].

### Outcomes

The primary end points of this pilot study are the feasibility and usability of FibroX-enabled care within diabetes-focused clinical workflows. Feasibility will be assessed using recruitment and study completion rates, with prespecified targets of enrolling at least 70% of eligible clinicians and achieving a completion rate of at least 85% among enrolled participants. Completion will be defined as participants completing both simulation periods, including all assigned cases and postperiod surveys. Participants who do not complete both periods or required assessments will be classified as incomplete. Usability will be evaluated using the System Usability Scale (SUS), with a prespecified success threshold of 70 or higher.

Secondary end points include workflow efficiency and preliminary diagnostic performance. Workflow efficiency will be assessed by median time per case, with an anticipated threshold of 3.5 minutes or less. Preliminary diagnostic performance outcomes will include sensitivity, specificity, and the area under the receiver operating characteristic curve for the detection of MASLD-associated advanced fibrosis. Calibration will be examined using intercept, slope, and graphical assessment of predicted versus observed risk.

Exploratory end points will include clinician trust and cognitive workload, assessed using the AI trust scale and National Aeronautics and Space Administration Task Load Index (NASA-TLX), respectively, as well as implementation-related outcomes, including the proportion of guideline-concordant referrals and net reclassification improvement. Qualitative debrief interviews will further explore clinician understanding, interpretability, and perceived usefulness of the SHAP-based explanations, including how these outputs influence decision-making and trust in the model. These measures are intended to provide pilot estimates and inform the design, end point selection, and sample size considerations for a future definitive trial.

At a high level, this pilot will be considered successful if predefined feasibility and usability thresholds are met, including adequate recruitment and completion rates and a SUS score of 70 or higher, alongside signals of acceptable workflow efficiency and improved or comparable diagnostic performance

relative to usual care. These criteria are intended to determine whether progression to a larger, real-world trial is justified.

### Participant Timeline and Study Schedule

This health care provider-level crossover pilot is planned over a 12-month period. Months 0 to 2 will be used for finalizing the platform and completing quality assurance of the simulated case library. The first study period will then be conducted during months 2 to 4, followed by a 1-week washout period to reduce potential carryover effects. The second study period will take place during months 4 to 6. After completion of both periods, participants will complete postperiod surveys assessing usability, trust, and cognitive workload, followed by brief semistructured debrief interviews. The remaining months will be dedicated to data cleaning, analysis, qualitative review, and final reporting.

### Sample Size

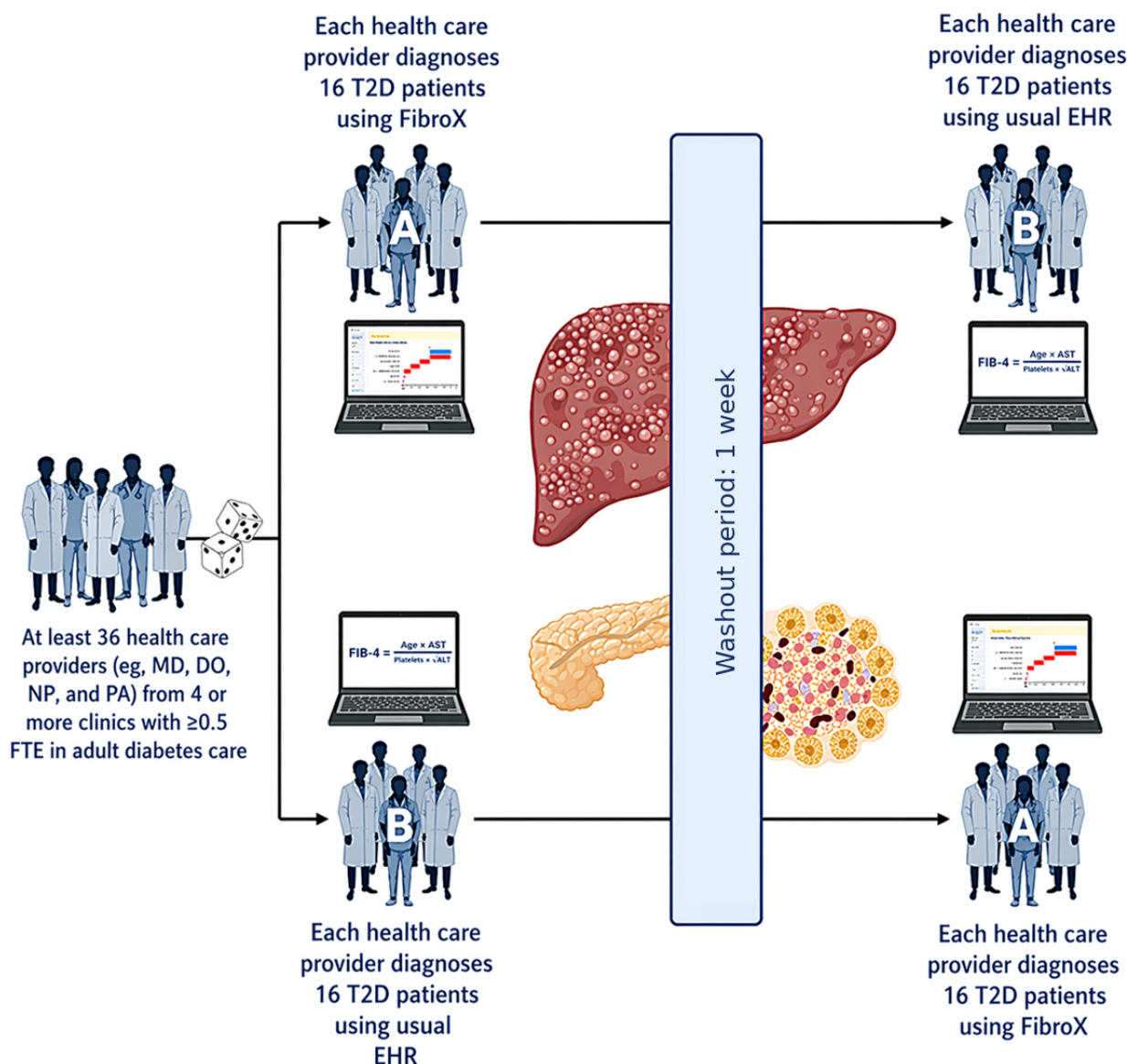
At least 36 health care providers (32 cases/health care provider) yield 1150 to 1280 paired decisions. Under pilot assumptions

(accuracy 0.60 usual care vs 0.75 FibroX), McNemar test provides >80% power to detect a 15-point improvement at  $\alpha=.05$  after accounting for health care provider-level clustering via a conservative design effect. Precision estimates for feasibility and usability metrics are expected to be narrow enough to inform a subsequent multicenter trial.

### Randomization, Allocation Concealment, and Implementation

Health care providers will be randomized 1:1 to AB (FibroX then usual care) or BA (usual care then FibroX) using a computer-generated schedule prepared by a statistician not involved in recruitment (Figure 2). Allocation will be concealed via a secure, centralized randomization module until assignment. The research coordinator enrolls participants and implements sequence assignments.

**Figure 2.** FibroX health care provider-level crossover randomized controlled trial pilot design. DO: doctor of osteopathic medicine; EHR: electronic health record; FTE: full-time equivalent; MD: doctor of medicine; NP: nurse practitioner; PA: physician assistant; T2D: type 2 diabetes.



## Blinding

Blinding of health care providers is not feasible due to visible tool features. Outcome assessors and data analysts are masked to sequence assignment; datasets use anonymized identifiers.

## Data Collection Methods

The platform automatically logs case navigation, decisions, timestamps, and use of FibroX features and overrides. Clinicians will not receive case-level performance feedback during either simulation period. Feedback, if provided, will be deferred until completion of the study period to minimize within-period learning effects. Surveys (SUS, AI trust, and NASA-TLX) are administered electronically at the end of each period. Qualitative debriefs are audio-recorded and transcribed for thematic analysis. Completeness checks and range validations occur at ingestion.

## Statistical Analysis

All analyses will follow an intention-to-treat principle, including all health care provider decisions regardless of sequence assignment. The primary comparison of diagnostic accuracy between FibroX-enabled care and usual care will use paired analyses to account for the crossover design. Sensitivity and specificity for detecting MASLD-associated advanced fibrosis will be compared using McNemar test for paired proportions. Overall discrimination will be assessed by calculating the area under the receiver operating characteristic curve for each condition, and calibration will be evaluated through intercept, slope, and calibration plots to determine agreement between predicted probabilities and observed outcomes.

To account for clustering at the health care provider level and the crossover design, mixed-effects logistic regression models will be used, specifying health care provider as a random effect and study condition, study period, and sequence assignment as fixed effects. Sensitivity analyses will also assess whether diagnostic performance differs by sequence assignment or period, which may suggest residual carryover or learning effects. These models will also adjust for case-level covariates such as age, BMI, and HbA<sub>1c</sub> to explore potential confounding and heterogeneity. Secondary analyses will include net reclassification improvement to quantify whether FibroX improves assignment of cases into clinically appropriate triage categories compared with usual care. Bootstrap resampling will be used to generate CIs for net reclassification improvement and calibration metrics, ensuring robust estimation under small-sample conditions.

Feasibility outcomes, including recruitment and completion rates, will be summarized descriptively with proportions and 95% CIs. Usability and trust scores from the SUS and AI trust scale will be analyzed using paired *t* tests or Wilcoxon signed-rank tests, depending on distributional assumptions. Cognitive workload scores from NASA-TLX will be compared similarly. Subgroup analyses will explore whether diagnostic performance varies across diabetes-related characteristics such as HbA<sub>1c</sub>, BMI, and duration of diabetes, using interaction terms in mixed-effects models. Missing data are expected to be minimal; if more than 5% of survey data are missing, multiple imputation under a missing-at-random assumption will be

applied, with sensitivity analyses conducted using complete-case data. All statistical tests will be 2-sided with an  $\alpha$  level of .05, and analyses will be performed using the R (R Foundation for Statistical Computing) or Python (Python Software Foundation) statistical packages.

To aid interpretation of the crossover design, sequence, period, and potential carryover effects will be explicitly examined. Sequence effects (AB vs BA) will assess whether the order of exposure to FibroX influences outcomes. Period effects will evaluate whether outcomes differ between the first and second study periods, independent of the intervention, which may reflect learning or temporal effects. Potential carryover effects will be explored by assessing interaction terms between intervention and study period and by comparing outcomes across sequences; a significant difference by sequence or interaction with period would suggest residual carryover. If meaningful carryover effects are detected, sensitivity analyses restricted to first-period data will be performed.

## Monitoring, Harms, and Auditing

Given simulated, minimal-risk procedures, no independent monitoring board will be convened. The research coordinator will track enrollment, completion, protocol adherence, and technical issues weekly. Anticipated risks are limited to time burden and performance discomfort; scheduling flexibility, planned breaks, and aggregated reporting mitigate these risks.

## Ethical Considerations

Institutional review board (IRB) approval has been obtained from the Yale University IRB (HIC 2000027433). Electronic informed consent will be obtained from all health care providers prior to participation. In addition, protocol amendments will be documented, versioned, and communicated to the IRB and trial registry before implementation when required. Participants will not receive financial compensation for participation. All data will be handled in accordance with institutional privacy and confidentiality policies, with secure storage on encrypted servers and restricted access to authorized study personnel only.

All data reside on Health Insurance Portability and Accountability Act (HIPAA)-compliant, access-controlled servers with encryption at rest and in transit, role-based permissions, and audit trails. A master linkage file is stored separately. Only authorized staff have access to identifiable enrollment information; analysis datasets are deidentified. Data retention follows institutional policy and sponsor requirements.

## Study Status and Timeline

This manuscript describes a prospective study protocol. At the time of submission, participant recruitment and data collection had not yet commenced. The trial was registered with ClinicalTrials.gov (NCT07305324). Additional study status details and anticipated milestones are summarized in the Results section.

## Reporting and Dissemination

This protocol has been developed in accordance with the SPIRIT (Standard Protocol Items: Recommendations for Interventional Trials) 2025 statement and the SPIRIT-AI extension for clinical trial protocols involving AI intervention [24,25]. The study

schedule is presented in [Table 1](#), in accordance with the SPIRIT framework. The completed SPIRIT checklist is provided as [Multimedia Appendix 1](#). Upon completion of the study, findings will be submitted for presentation at a relevant professional society meeting and for publication in a peer-reviewed journal. The final results manuscript will adhere to CONSORT (Consolidated Standards of Reporting Trials) guidelines [26,27]. In addition, lay summaries will be prepared for participating

sites and disseminated through professional networks to facilitate knowledge translation and support future implementation. Deidentified datasets and analysis code will be made available upon reasonable request following publication and appropriate data-sharing agreements. Findings will be disseminated via conferences, a peer-reviewed manuscript, and lay summaries for participating sites.

**Table 1.** Study schedule of enrollment, interventions, and outcome assessments (SPIRIT [Standard Protocol Items: Recommendations for Interventional Trials] framework).

Domain and procedure	Enrollment	Allocation	Intervention period 1	Washout	Intervention period 2	Closeout
	T-2 weeks to T0 <sup>a</sup>	T0	T1 <sup>b</sup> (1-2 wk)	T2 <sup>c</sup> (1 wk)	T3 <sup>d</sup> (1-2 wk)	T4 <sup>e</sup>
<b>Enrollment</b>						
Eligibility screen and recruitment	✓					
Informed consent	✓					
Orientation and technical competency check	✓					
<b>Allocation</b>						
Randomization to sequence (AB <sup>f</sup> or BA <sup>g</sup> )		✓				
<b>Intervention</b>						
FibroX-enabled care (condition A)			✓ <sup>h</sup>		✓ <sup>i</sup>	
Usual care (condition B)			✓ <sup>i</sup>		✓ <sup>h</sup>	
<b>Assessments</b>						
<b>Feasibility outcomes</b>						
Decision time (workflow efficiency)			✓		✓	
Usability (SUS <sup>j</sup> score)			✓		✓	
Trust in AI <sup>k</sup> (AI trust scale)			✓		✓	
Cognitive workload (NASA-TLX <sup>l</sup> )			✓		✓	
<b>Effectiveness outcome</b>						
Diagnostic accuracy (sensitivity, specificity, and AUROC <sup>m</sup> )			✓		✓	
<b>Secondary outcomes</b>						
Referral rates, confidence, and override rates			✓		✓	
Qualitative debrief interview (barriers and facilitators)						✓
Final data verification and feedback						✓

<sup>a</sup>T0: baseline or randomization.

<sup>b</sup>T1: first intervention period.

<sup>c</sup>T2: washout.

<sup>d</sup>T3: second intervention period.

<sup>e</sup>T4: study closeout.

<sup>f</sup>Sequence AB: FibroX-enabled care in period 1 followed by usual care in period 2.

<sup>g</sup>Sequence BA: usual care in period 1 followed by FibroX-enabled care in period 2.

<sup>h</sup>In Sequence AB.

<sup>i</sup>In Sequence BA.

<sup>j</sup>SUS: System Usability Scale.

<sup>k</sup>AI: artificial intelligence.

<sup>l</sup>NASA-TLX: National Aeronautics and Space Administration Task Load Index.

<sup>m</sup>AUROC: area under the receiver operating characteristic curve.

## Results

IRB approval has been obtained. At the time of manuscript submission, participant recruitment had not yet commenced. Preparatory activities, including platform finalization, simulation interface testing, and case library validation, were ongoing in advance of the anticipated study start date of June 15, 2026. Primary completion is anticipated on May 15, 2027, with overall study completion anticipated on June 15, 2027.

## Discussion

### Anticipated Findings

This pilot trial addresses a critical gap in diabetes care: the underdiagnosis of MASLD-associated advanced liver fibrosis despite its strong association with adverse hepatic and cardiovascular outcomes. Current noninvasive tools underperform in T2D populations, and workflow barriers further limit guideline-concordant screening. Because FibroX is not currently used in routine clinical practice, this study evaluates its feasibility, usability, and preliminary diagnostic effectiveness within simulated clinical workflows before progression to real-world implementation. A health care provider-level randomized crossover simulation design was chosen to ensure each clinician serves as their own control, improving internal validity and reducing between-health care provider variability while enabling rigorous paired analyses. Simulation with deidentified or synthetic cases minimizes risk and isolates decision performance from external clinical confounders while preserving the EHR-like context needed to assess workflow integration. This approach allows preliminary evaluation of feasibility, usability, and diagnostic performance in a controlled setting before progression to real-world clinical testing. In addition to improving risk estimation, the inclusion of SHAP-based explanations is designed to make model outputs more interpretable and actionable, potentially supporting more consistent and guideline-aligned decision-making compared with usual care.

### Comparison to Prior Work

Prior studies of machine learning in MASLD and fibrosis risk stratification have largely focused on retrospective model development and diagnostic performance, with comparatively

less attention to clinician-facing usability, workflow integration, and implementation [19,21,28]. Explainable approaches such as SHAP have been proposed to improve interpretability, but prospective evaluation of how such tools influence clinical decision-making remains limited [29,30]. In this context, this pilot study extends prior work by moving beyond retrospective validation to assess FibroX as a clinician-facing decision support tool within a simulated workflow environment.

### Strengths and Limitations

This study has several strengths, including its randomized crossover design, within-clinician comparison, standardized simulated cases, and explicit assessment of feasibility, usability, and preliminary diagnostic performance. At the same time, it has limitations inherent to simulation-based designs, including the absence of real-world workflow pressures, competing clinical demands, and direct patient interaction. These features make the study well-suited for an initial pilot evaluation, but clinician behavior in simulated scenarios may not fully reflect decision-making in routine practice. Future research will therefore need to validate these findings in real clinical settings, assess longer-term implementation and outcomes, and examine scalability across diverse health systems, including resource-limited environments where MASLD burden is high.

### Future Work

If FibroX demonstrates acceptable feasibility, usability, and preliminary diagnostic performance in this pilot setting, the next step will be a larger pragmatic study in real-world clinical practice to evaluate workflow impact, referral patterns, and diagnostic effectiveness under routine care conditions. Such work should also assess generalizability across diverse practice settings, including community clinics and resource-limited environments, examine longer-term implementation outcomes, and further evaluate whether SHAP-based explanations improve interpretability and clinician trust in routine care.

### Conclusions

In conclusion, this pilot will provide early evidence on the feasibility, usability, and preliminary clinical utility of FibroX within simulated diabetes care workflows. The findings will help determine whether further real-world evaluation of this explainable AI approach is warranted for MASLD risk stratification in people with T2D.

## Acknowledgments

Generative artificial intelligence tools were used for minor language editing and grammar refinement. All content was reviewed and approved by the authors, who take full responsibility for the accuracy and integrity of the manuscript.

## Funding

This study did not receive any specific funding.

## Data Availability

As this study is a protocol involving simulated clinical cases, no real patient-level data will be generated. Deidentified study data, including aggregated decision outcomes and analysis code, will be made available from the corresponding author upon reasonable request following study completion and publication, subject to institutional policies and data use agreements. Any materials necessary to reproduce the simulation environment or analysis framework may also be shared upon request where feasible.

## Authors' Contributions

BN conceptualized the study and supervised the project. BN and USK contributed to the study design, methodology, and drafting of the manuscript. Both authors reviewed, edited, and approved the final version of the manuscript.

## Conflicts of Interest

BN and USK were involved in the development of the FibroX model evaluated in this study.

## Multimedia Appendix 1

SPIRIT checklist.

[\[DOCX File, 33 KB-Multimedia Appendix 1\]](#)

## References

1. Byrne CD, Targher G, Buchanan RM. Type 2 diabetes, prediabetes, and MASLD: who and when to screen and how to treat? *Lancet Diabetes Endocrinol*. Aug 2025;13(8):645-646. [doi: [10.1016/S2213-8587\(25\)00189-5](https://doi.org/10.1016/S2213-8587(25)00189-5)] [Medline: [40544852](https://pubmed.ncbi.nlm.nih.gov/40544852/)]
2. Nwoguh C, Byrne CD, Reinson T, Buchanan RM. Screening for MASLD in patients with type 2 diabetes: is an early diagnosis a good diagnosis? *Expert Rev Gastroenterol Hepatol*. Sep 19, 2025;19(9):941-951. [doi: [10.1080/17474124.2025.2546597](https://doi.org/10.1080/17474124.2025.2546597)] [Medline: [40824753](https://pubmed.ncbi.nlm.nih.gov/40824753/)]
3. Cusi K, Abdelmalek MF, Apovian CM, Balapattabi K, Bannuru RR, Barb D, et al. Metabolic Dysfunction-Associated Steatotic Liver Disease (MASLD) in people with diabetes: the need for screening and early intervention. A consensus report of the American Diabetes Association. *Diabetes Care*. Jul 01, 2025;48(7):1057-1082. [doi: [10.2337/dci24-0094](https://doi.org/10.2337/dci24-0094)] [Medline: [40434108](https://pubmed.ncbi.nlm.nih.gov/40434108/)]
4. Balkhed W, Bergram M, Iredahl F, Holmberg M, Edin C, Carlhäll CJ, et al. Evaluating the prevalence and severity of metabolic dysfunction-associated steatotic liver disease in patients with type 2 diabetes mellitus in primary care. *J Intern Med*. Sep 16, 2025;298(3):173-187. [FREE Full text] [doi: [10.1111/joim.20103](https://doi.org/10.1111/joim.20103)] [Medline: [40518766](https://pubmed.ncbi.nlm.nih.gov/40518766/)]
5. Nwosu C, Batakana S, Vissoci J, Vaca S, Lim S, Smith E, et al. Identifying the needs and barriers to patient-family education to design educational interventions that will improve neurosurgery patient outcomes in Mulago Hospital, Uganda. *Ann Glob Health*. 2017;83(1):123-124. [doi: [10.1016/j.aogh.2017.03.275](https://doi.org/10.1016/j.aogh.2017.03.275)]
6. Hudson D, Afzaal T, Bualbanat H, AlRamdan R, Howarth N, Parthasarathy P, et al. Modernizing metabolic dysfunction-associated steatotic liver disease diagnostics: the progressive shift from liver biopsy to noninvasive techniques. *Ther Adv Gastroenterol*. Nov 14, 2024;17:17562848241276334. [FREE Full text] [doi: [10.1177/17562848241276334](https://doi.org/10.1177/17562848241276334)] [Medline: [39553445](https://pubmed.ncbi.nlm.nih.gov/39553445/)]
7. Frączek J, Sowa A, Agopsowicz P, Migacz M, Dylińska-Kala K, Holecki M. Non-invasive tests as a replacement for liver biopsy in the assessment of MASLD. *Medicina (Kaunas)*. Apr 16, 2025;61(4):736. [FREE Full text] [doi: [10.3390/medicina61040736](https://doi.org/10.3390/medicina61040736)] [Medline: [40283027](https://pubmed.ncbi.nlm.nih.gov/40283027/)]
8. Dawod S, Brown K. Non-invasive testing in metabolic dysfunction-associated steatotic liver disease. *Front Med (Lausanne)*. Nov 13, 2024;11:1499013. [FREE Full text] [doi: [10.3389/fmed.2024.1499013](https://doi.org/10.3389/fmed.2024.1499013)] [Medline: [39606621](https://pubmed.ncbi.nlm.nih.gov/39606621/)]
9. Yoon JH, Lee JM, Joo I, Lee ES, Sohn JY, Jang SK, et al. Hepatic fibrosis: prospective comparison of MR elastography and US shear-wave elastography for evaluation. *Radiology*. Dec 2014;273(3):772-782. [doi: [10.1148/radiol.14132000](https://doi.org/10.1148/radiol.14132000)] [Medline: [25007047](https://pubmed.ncbi.nlm.nih.gov/25007047/)]
10. Wong VW, Vergniol J, Wong GL, Foucher J, Chan HL, Le Bail B, et al. Diagnosis of fibrosis and cirrhosis using liver stiffness measurement in nonalcoholic fatty liver disease. *Hepatology*. Feb 2010;51(2):454-462. [doi: [10.1002/hep.23312](https://doi.org/10.1002/hep.23312)] [Medline: [20101745](https://pubmed.ncbi.nlm.nih.gov/20101745/)]
11. Boursier J, de Ledinghen V, Zarski JP, Fouchard-Hubert I, Gallois Y, Oberti F, et al. Comparison of eight diagnostic algorithms for liver fibrosis in hepatitis C: new algorithms are more precise and entirely noninvasive. *Hepatology*. Jan 2012;55(1):58-67. [doi: [10.1002/hep.24654](https://doi.org/10.1002/hep.24654)] [Medline: [21898504](https://pubmed.ncbi.nlm.nih.gov/21898504/)]
12. Meng F, Zheng Y, Zhang Q, Mu X, Xu X, Zhang H, et al. Noninvasive evaluation of liver fibrosis using real-time tissue elastography and transient elastography (FibroScan). *J Ultrasound Med*. Mar 2015;34(3):403-410. [doi: [10.7863/ultra.34.3.403](https://doi.org/10.7863/ultra.34.3.403)] [Medline: [25715361](https://pubmed.ncbi.nlm.nih.gov/25715361/)]
13. Lam R, Jain D, Deng Y, Acharya E, Lim JK. Advanced liver fibrosis predicts liver outcomes in biopsy-proven metabolic dysfunction-associated steatotic liver disease: a U.S.-based single-center retrospective cohort study. *J Clin Transl Hepatol*. Dec 28, 2024;12(12):988-996. [FREE Full text] [doi: [10.14218/JCTH.2024.00189](https://doi.org/10.14218/JCTH.2024.00189)] [Medline: [39649030](https://pubmed.ncbi.nlm.nih.gov/39649030/)]
14. Ajmera V, Tesfai K, Sandoval E, Lopez S, Cervantes V, Madamba E, et al. Validation of AGA clinical care pathway and AASLD practice guidance for nonalcoholic fatty liver disease in a prospective cohort of patients with type 2 diabetes. *Hepatology*. May 01, 2024;79(5):1098-1106. [doi: [10.1097/HEP.0000000000000635](https://doi.org/10.1097/HEP.0000000000000635)] [Medline: [37862551](https://pubmed.ncbi.nlm.nih.gov/37862551/)]
15. Noguerol Álvarez M, Valer López Fando MP, Torrijos Bravo C, Gómez Ortiz MC, Piqueras Alcohol B, Guardiola Arévalo A, et al. Screening for advanced liver disease incorporating the use of transitional elastography in primary care. *Gastroenterología y Hepatología (English Edition)*. Mar 2025;48(3):502242. [doi: [10.1016/j.gastre.2025.502242](https://doi.org/10.1016/j.gastre.2025.502242)]

16. Decharatanachart P, Chaiteerakij R, Tiyyarattanachai T, Treeprasertsuk S. Application of artificial intelligence in non-alcoholic fatty liver disease and liver fibrosis: a systematic review and meta-analysis. *Ther Adv Gastroenterol*. Dec 21, 2021;14:17562848211062807. [FREE Full text] [doi: [10.1177/17562848211062807](https://doi.org/10.1177/17562848211062807)] [Medline: [34987607](https://pubmed.ncbi.nlm.nih.gov/34987607/)]
17. Njei B, Osta E, Ilagan-Ying YC, Boateng S, Al-Ajlouni YA, Al Ta'ani O, et al. Fibrox: an explainable ai model for accurate prediction of advanced liver fibrosis and cardiovascular mortality in MASLD. *Gastroenterology*. May 2025;169(1):S-131-S-132. [doi: [10.1016/S0016-5085\(25\)01246-6](https://doi.org/10.1016/S0016-5085(25)01246-6)]
18. Njei B, Osta E, Njei N, Al-Ajlouni YA, Lim JK. An explainable machine learning model for prediction of high-risk nonalcoholic steatohepatitis. *Sci Rep*. Apr 13, 2024;14(1):8589. [FREE Full text] [doi: [10.1038/s41598-024-59183-4](https://doi.org/10.1038/s41598-024-59183-4)] [Medline: [38615137](https://pubmed.ncbi.nlm.nih.gov/38615137/)]
19. Njei B, Al-Ajlouni YA, Lemos SY, Ugwendum D, Njei N, Al Ta'ani O, et al. AI-based models for risk prediction in MASLD: a systematic review. *Dig Dis Sci*. Nov 13, 2025;169(1):S-131-S-132. [doi: [10.1007/s10620-025-09499-6](https://doi.org/10.1007/s10620-025-09499-6)] [Medline: [41231418](https://pubmed.ncbi.nlm.nih.gov/41231418/)]
20. Payrovnaziri SN, Chen Z, Rengifo-Moreno P, Miller T, Bian J, Chen JH, et al. Explainable artificial intelligence models using real-world electronic health record data: a systematic scoping review. *J Am Med Inform Assoc*. Jul 01, 2020;27(7):1173-1185. [FREE Full text] [doi: [10.1093/jamia/ocaa053](https://doi.org/10.1093/jamia/ocaa053)] [Medline: [32417928](https://pubmed.ncbi.nlm.nih.gov/32417928/)]
21. Dabbah S, Mishani I, Davidov Y, Ben Ari Z. Implementation of machine learning algorithms to screen for advanced liver fibrosis in metabolic dysfunction-associated steatotic liver disease: an in-depth explanatory analysis. *Digestion*. Oct 25, 2025;106(3):189-202. [FREE Full text] [doi: [10.1159/000542241](https://doi.org/10.1159/000542241)] [Medline: [39462487](https://pubmed.ncbi.nlm.nih.gov/39462487/)]
22. Abbas Q, Jeong W, Lee SW. Explainable AI in clinical decision support systems: a meta-analysis of methods, applications, and usability challenges. *Healthcare (Basel)*. Aug 29, 2025;13(17):2154. [FREE Full text] [doi: [10.3390/healthcare13172154](https://doi.org/10.3390/healthcare13172154)] [Medline: [40941506](https://pubmed.ncbi.nlm.nih.gov/40941506/)]
23. 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*. Jun 2006;43(6):1317-1325. [doi: [10.1002/hep.21178](https://doi.org/10.1002/hep.21178)] [Medline: [16729309](https://pubmed.ncbi.nlm.nih.gov/16729309/)]
24. Chan AW, Boutron I, Hopewell S, Moher D, Schulz KF, Collins GS, et al. SPIRIT 2025 statement: updated guideline for protocols of randomised trials. *Lancet*. May 2025;405(10491):e19-e27. [doi: [10.1016/S0140-6736\(25\)00770-6](https://doi.org/10.1016/S0140-6736(25)00770-6)]
25. Rivera SC, Liu X, Chan A, Denniston AK, Calvert MJ, SPIRIT-AICONSORT-AI Working Group. Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI extension. *BMJ*. Sep 09, 2020;370:m3210. [FREE Full text] [doi: [10.1136/bmj.m3210](https://doi.org/10.1136/bmj.m3210)] [Medline: [32907797](https://pubmed.ncbi.nlm.nih.gov/32907797/)]
26. Hopewell S, Chan A, Collins GS, Hróbjartsson A, Moher D, Schulz KF, et al. CONSORT 2025 statement: updated guideline for reporting randomised trials. *Lancet*. Apr 14, 2025;405(10489):1633-1640. [doi: [10.1016/S0140-6736\(25\)00672-5](https://doi.org/10.1016/S0140-6736(25)00672-5)] [Medline: [40245901](https://pubmed.ncbi.nlm.nih.gov/40245901/)]
27. Ibrahim H, Liu X, Rivera SC, Moher D, Chan A, Sydes MR, et al. Reporting guidelines for clinical trials of artificial intelligence interventions: the SPIRIT-AI and CONSORT-AI guidelines. *Trials*. Jan 06, 2021;22(1):11. [FREE Full text] [doi: [10.1186/s13063-020-04951-6](https://doi.org/10.1186/s13063-020-04951-6)] [Medline: [33407780](https://pubmed.ncbi.nlm.nih.gov/33407780/)]
28. McTeer M, Applegate D, Mesenbrink P, Ratzu V, Schattenberg JM, Bugianesi E, et al. Machine learning approaches to enhance diagnosis and staging of patients with MASLD using routinely available clinical information. *PLoS One*. Feb 29, 2024;19(2):e0299487. [FREE Full text] [doi: [10.1371/journal.pone.0299487](https://doi.org/10.1371/journal.pone.0299487)] [Medline: [38421999](https://pubmed.ncbi.nlm.nih.gov/38421999/)]
29. Roychowdhury S, Lanfranchi V, Mazumdar S. Evaluating explanation performance for clinical decision support systems for non-imaging data: a systematic literature review. *Comput Biol Med*. Oct 2025;197(Pt A):110944. [doi: [10.1016/j.compbimed.2025.110944](https://doi.org/10.1016/j.compbimed.2025.110944)] [Medline: [40857813](https://pubmed.ncbi.nlm.nih.gov/40857813/)]
30. Van Dort BA, Engelsma T, Medlock S, Dusseljee-Peute L. User-centered methods in explainable AI development for hospital clinical decision support: a scoping review. *Stud Health Technol Inform*. May 12, 2025;326:17-21. [doi: [10.3233/SHTI250228](https://doi.org/10.3233/SHTI250228)] [Medline: [40357594](https://pubmed.ncbi.nlm.nih.gov/40357594/)]

## Abbreviations

- AI:** artificial intelligence
- ALT:** alanine aminotransferase
- AST:** aspartate aminotransferase
- CONSORT:** Consolidated Standards of Reporting Trials
- EHR:** electronic health record
- FIB-4:** Fibrosis-4 Index
- HbA<sub>1c</sub>:** hemoglobin A<sub>1c</sub>
- HIPAA:** Health Insurance Portability and Accountability Act
- IRB:** institutional review board
- MASLD:** metabolic dysfunction-associated steatotic liver disease
- NASA-TLX:** National Aeronautics and Space Administration Task Load Index
- SHAP:** Shapley Additive Explanations

**SPIRIT:** Standard Protocol Items: Recommendations for Interventional Trials

**SUS:** System Usability Scale

**T2D:** type 2 diabetes

**VCTE:** Vibration-Controlled Transient Elastography

*Edited by A Schwartz; submitted 28.Dec.2025; peer-reviewed by X Liang, G Toda; comments to author 23.Jan.2026; revised version received 12.Apr.2026; accepted 13.Apr.2026; published 21.May.2026*

*Please cite as:*

*Njei B, Kanmounye US*

*An Explainable AI Tool (FibroX) for Detecting Advanced Liver Fibrosis in Adults With Type 2 Diabetes: Protocol for a Pilot Crossover Trial*

*JMIR Res Protoc 2026;15:e90456*

*URL: <https://www.researchprotocols.org/2026/1/e90456>*

*doi: [10.2196/90456](https://doi.org/10.2196/90456)*

*PMID:*

©Basile Njei, Ulrick Sidney Kanmounye. Originally published in JMIR Research Protocols (<https://www.researchprotocols.org>), 21.May.2026. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Research Protocols, is properly cited. The complete bibliographic information, a link to the original publication on <https://www.researchprotocols.org>, as well as this copyright and license information must be included.