Title: Physiological foundation modeling for subclinical disease assessment: a prospective pilot

Authors: Yuan, William, PhD[1]; Xu, Shiwei, PhD[1]; Dionisi, Sara, PhD[2]; von Herrmann, Katharine, PhD[1,2]; Frederick-Dyer, Kate, MD[3]; Cheung-Flynn, Joyce, PhD[4]; Kennedy, Caleb, PhD[1*]; Flynn, Charles R, PhD[5*]; Lipnick, Scott, PhD[1,2*]

*co-senior author

Affiliations: 1) Etiome Inc, Cambridge, MA; 2) Flagship Pioneering, Cambridge, MA; 3) Department of Radiology and Radiological Sciences, Vanderbilt University Medical Center, Nashville, TN, USA**,** 4) Department of Vascular Surgery, Vanderbilt University Medical Center, Nashville, TN, USA, 5) Department of Surgery, Vanderbilt University Medical Center, Nashville, TN, USA

Author Contributions:

CK had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Conceptualization: W.Y., S.L., R.F., C.K.; **Methodology:** W.Y., S.X.; **Software:** W.Y., S.X.; **Validation:** W.Y., S.X.; **Investigation:** W.Y., S.X., S.D., K.F.-D., J.C.-F., R.F.; **Resources:** W.Y., J.C.-F., R.F.; **Data Curation:** W.Y., S.X., J.C.-F., R.F.; **Writing:** all authors; **Supervision:** S.L., R.F., C.K.; **Administration:** S.L., R.F., C.K.

COI Disclosures-: W.Y, S.X, and C.K are employees of Etiome Inc. S.D, K.VH, and S.L are employees of Flagship Pioneering. C.R.F, J.C.-F, and K.F.-D are employed by Vanderbilt University Medical Center.

Funding and Support: This study was funded in full by Etiome Inc.

Data Sharing Statement: The UK Biobank is available for use at https://www.ukbiobank.ac.uk by submitting a data request. The FLINT and NAFLD Adult databases are available for request at https://repository.niddk.nih.gov. The specific data from these studies used in this current work were used under license and so are not available. NHANES is freely available for use at https://www.cdc.gov/nchs/nhanes/index.html. The data from the Gifu, Kaohsiung, and Iksan cohorts are freely available at their respective journals. The Dandelion cohort was used under license and so is not available. The remaining data and analytic methods used in this study will remain the confidential property of the study sponsors. Data may be made available to other researchers upon request following permission, investigator support and following a signed data access agreement.

Word Count: 2955

Abbreviations: MASLD, metabolic dysfunction-associated steatotic liver disease; PDFF, proton density fat fraction.

Corresponding Author: William Yuan; wyuan@etiome.bio

Abstract (284):

Clinical studies across disease areas struggle to locate the right patients, in part because many remain undiagnosed or lack relevant subclinical disease labels. We created an AI framework that converts routine health-record data into physiology-based “Bioprofiles,” enabling faster, more precise identification of candidates for prospective research. We aimed to evaluate whether Bioprofile-guided recruitment could reduce the number of individuals who must be screened and enable more precise, physiology-informed targeting of candidates for steatotic liver disease studies. We trained Bioprofile patient representations from routinely collected health data including demographics, medical history, lab results, and lifestyle factors, and fine-tuned against multiple endpoints including MASLD severity, evaluated using Proton density fat fraction (PDFF), an imaging-based metric of steatosis severity. We trained Bioprofiles using 1 million subjects from the UK Biobank and other cohorts and recruited 31 subjects from Vanderbilt University Medical Center for study visits based on Bioprofile nominations. Bioprofile models achieved a Spearman coefficient of 0.646 against PDFF, strongly outperforming existing foundation models (r = 0.361) and clinical risk scores (r = 0.516-0.545). Simulations of the Bioprofile patient nomination pipeline suggest it requires half as many subjects needed to screen compared to existing methods depending on task. These findings were validated in a prospective pilot at VUMC, where 31 Bioprofile-nominated subjects were recruited to undergo liver imaging. Bioprofile predictions aligned strongly with study data (r = 0.740). Bioprofiles were further validated against 5 global cohort studies for non-PDFF MASLD endpoints. AI-based profiling of patient physiology can surface individuals who harbor subclinical or otherwise unrecognized disease signatures. This approach can facilitate subject nomination for human studies, reduce screening failures, and improve trial quality. Bioprofiles may have additional utility in decision support applications in precision medicine and AI-augmented healthcare.

Introduction

Metabolic dysfunction-associated steatotic liver disease (MASLD; formerly non-alcoholic fatty liver disease, NAFLD[1]) is highly prevalent yet remains markedly underdiagnosed in clinical practice[2]. The condition, defined by liver fat accumulation (hepatic steatosis) in the presence of metabolic risk factors, affects over a quarter of adults worldwide[3]. However, many individuals with MASLD are never diagnosed, owing to its often subclinical course and the limited use of definitive diagnostics in routine care[4–6]. This under-recognition poses a major challenge for patient recruitment in clinical studies, as potentially eligible participants may not even know they have steatotic liver disease. Even when MASLD is suspected, confirming steatosis severity typically requires imaging or biopsy, introducing logistical or financial hurdles in human trials[7,8]. MRI-based proton density fat fraction (PDFF) measurements are considered a gold-standard for noninvasive quantification of liver fat and are commonly used as an inclusion criterion for clinical trials. Changes in PDFF are also used as an endpoint of treatment efficacy for steatosis reduction[9–11] However, MRI is expensive, time-intensive, and not accessible at scale[12,13,14].

There is a clear need for steatosis metrics that are both non-invasive and scalable to flag high-risk patients. Simple clinical scores[15,16] exist to estimate MASLD probability from routine labs and demographics but offer only moderate accuracy and were not designed to precisely quantify liver fat or guide trial enrollment. One recent study[17] developed an artificial intelligence

(AI) algorithm for identifying subjects based on electronic health records (EHRs) at particularly high risk and hypothesized that such an algorithm could support earlier screening and even facilitate recruitment for trials. However, this algorithm was only externally validated on retrospective real-world data where true disease labels were unavailable. Furthermore, human studies, particularly those focused on discovery efforts, may aim to identify subjects at various stages of disease, rather than just one. More broadly, foundation models, which utilize broad pre-training to enable prediction of multiple down-stream endpoints, have been applied to a wide set of healthcare applications, but liver specific applications and real-world validation have not been performed to our knowledge[18].

In this work, we introduce and validate Bioprofiles, a foundation model of human physiology trained on large, diverse cohorts, first applied in a cardiometabolic context to detect patterns linked to liver-fat accumulation. By integrating clinically accessible data (demographics, medical history, laboratory results, and lifestyle factors), Bioprofiles can estimate an individual's liver fat distribution without the need for new tests or imaging. Furthermore, steatosis identification can contribute to qualifying patients for glucagon-like peptide-1 receptor agonists (GLP1 RAs)[19]. Bioprofiles offer a non-invasive, widely applicable tool to nominate likely MASLD patients before any intervention or imaging is done, thereby broadening the pool of trial-eligible participants and enabling earlier therapeutic engagement. In this manuscript, we evaluate Bioprofiles in a prospective pilot for MASLD trial recruitment.

Methods

Data sources

Deidentified data from participants from the UK Biobank[20] (n = 500k), NHANES[21] (n = 90k) and an internal set of hospital-based cohort data (n = 460k) were included in this study. The UK Biobank was subdivided into pretraining (n = 450k) and fine-tuning (n = 43k) sets based on availability of liver imaging data. Summary statistics from a set of 45,484 subjects from Vanderbilt University Medical Center (VUMC) were also utilized for calibration.

Step 1: Model training

Clinically-accessible data elements including demographics, medical history, lab results, and lifestyle factors were extracted and manually harmonized across datasets. A foundation model was trained over the pre-training dataset to generate intermediate patient representations, which were fine-tuned against the label of interest using a diffusion model regression or classification head, depending on task definition (Supplementary Figure 1A). The fine-tuning set was itself split into non-overlapping training, test, and validation sets (65:10:25). The primary endpoint considered was PDFF, measured using magnetic resonance imaging (MRI), which measures the relative degree of contributing signal from fat versus water and is expressed as a percentage value[22].

To benchmark performance, we compared Bioprofile predictions to those generated by CLMBR, a recent foundation model that utilizes longitudinal electronic health records[23,24], the Framingham Steatosis Index (FSI)[15], and Hepatic Steatosis Index (HSI)[16]. We also employed two technical benchmarks: a gradient boosted regressor trained over the fine-tuning set, with

access to i) body mass index (BMI), alanine aminotransferase (ALT), and aspartate aminotransferase (AST) levels, and ii) all Bioprofile input features. Bioprofile performance was also evaluated relative to the full technical benchmark on a series of non-proton density fat fraction (PDFF) tasks (Supplementary Table 1).

Step 2: Targeted recruitment task:

The first recruitment task we considered was to identify subjects with PDFF > 10%. This corresponds to a common inclusion criterion for interventional trials, where sponsors are interested in ensuring that recruited subjects have a sufficient amount of steatosis[25,26]. In simulated studies and during the VUMC recruitment pilot, we aimed to minimize the number of subjects screened to recruit 12 subjects with PDFF > 10%. The specific recruitment targets for the study were derived from sample size (Supplement) and logistic study considerations.

Step 3: Even recruitment task:

The second recruitment task we considered was to identify a uniform number of subjects across PDFF, specifically PDFF between 0% and 5%, between 5% and 15%, and above 15%, corresponding to healthy, mild, and moderate degrees of steatosis[27]. These correspond to a discovery-related task where a sponsor is interested in obtaining a sufficient number of subjects across a transition between relative health and disease, for the purpose of developing or validating a diagnostic or biomarker[28]. In simulated studies and during the VUMC recruitment pilot, we aimed to minimize the number of subjects screened to recruit at least 9 subjects across all 3 bins.

Step 4: Cohort nomination algorithm

We developed an algorithm (Portfolio Selection, Supplementary Figure 1B) to nominate subjects for recruitment based on Bioprofile-generated probability distributions. Bioprofiles are first calculated for all subjects eligible for recruitment, which are used to generate conditional probability distributions over PDFF. We utilized linear programming to nominate the subset of subjects with a collective probability distribution that most closely resembles the goal distribution (all subjects with PDFF > 10% or even distribution of PDFF over 3 bins). Nominated subjects are measured and the true distribution of subjects is compared to the goal. If the recruitment goal is not yet reached, additional sub-cohorts are nominated and recruited. Notably, for the VUMC study, we were able to embed the targeted recruitment task within the even recruitment task, identifying a single population that simultaneously met both criteria (Supplementary Figure 1C).

Step 5: Study recruitment simulations

To evaluate the difficulty of obtaining cohorts with desired distributions, we simulated the ability of various models to guide subject recruitment studies across both tasks (targeted and even recruitment). Furthermore, we evaluated two methods for model-guided subject recruitment: standard sampling, and Portfolio Selection. For standard sampling in targeted recruitment, subjects were recruited in rank order based on their model-assessed probability of having PDFF > 10%. For standard sampling in even recruitment, subjects were recruited in batches of 5

based on rank order of model-assessed probability of membership in each of the three bins, with recruitment shifting across bins as recruitment goals are reached (adaptive sampling).

Subject screening and pre-recruitment

VUMC health records in the Epic data warehouse were eligible if they were between the ages of 21 and 70. Exclusion criteria for the fine-tuning cohort and VUMC recruitment cohort included pregnancy or breast-feeding, presence of implanted cardiac defibrillator or pacemaker, presence of cancer and/or ongoing cancer treatment, presence of HIV, history of alcohol consumption exceeding 14 drinks per week for men and greater than seven per week for women, illicit drug use, treatment with any investigational drug in the one month preceding study, mental conditions rendering an inability to understand the nature, scope and possible consequences of the study, inability to comply with the protocol (Supplementary Information). We also excluded subjects with existing diagnoses of MASLD from the VUMC recruitment cohort to evaluate the utility of our models on identifying undiagnosed MASLD. The study was approved by the Vanderbilt IRB (230833).

Subject recruitment and sampling

A recruitment cohort of 45,484 subjects was identified in the VUMC clinical data warehouse based on inclusion, exclusion, data density, and subject accessibility requirements (Supplementary Information). Bioprofiles and PDFF estimates were calculated on these subjects using features extracted from historical EHR. These nominations were made based on data present in their medical record (Pre-visit EHR Extract), but a full physical exam and lab series was performed at the study visit (Study Visit). Data on all cohorts are provided in Table 1. A CONSORT diagram of participant recruitment is provided in Supplementary Figure 2.

The MyHealth@Vanderbilt platform was used to consent and engage nominated subjects. Recruited subjects were invited to participate in a study visit consisting of body measurements, venipuncture, blood analyses, and liver magnetic resonance imaging with the purpose of measuring PDFF.

Subject schedule of study events

31 subjects who met inclusion and exclusion criteria were invited to the Vanderbilt Clinical Research Center after an overnight fast. A brief physical examination including measures of waist, hip, and chest circumference along with blood pressure were taken prior to a blood draw for a lipid panel, CBC, CMP and measurement of HbA1c. 2 subjects were found to be incompatible with MRI screening during the study visit. 29 subjects underwent MRI evaluation of the upper abdomen using the mDixon protocol. Results of estimated percentage of liver fat by MR method and Bioprofile-predicted liver fat content were correlated.

MRI examination of liver was done with a 3T MR7700 MR system (Philips Healthcare, Netherlands) with a quadrature body transmit coil used in combination with a 32-channel anterior receive coil array and 16-channel posterior receive coil array. Anatomic imaging of entire upper abdomen was performed with axial, free-breathing single-shot Turbo Spin Echo (TSE) sequence. A commercially available version of mDIXON sequence package was used to

acquire fat and water images of entire liver. Following were the mDIXON Quant imaging parameters: 3D T1-FFE sequence, FOR=400/350/231mm, acq. vox.dim.=3/3/6mm (over contiguous slices with 3mm overlap), 6-echoes: TE1/delta_TE=0.92/0.7 msec, TR=5.2 msec, Flip angle=3deg, SENSE =1.5,1.5 in two directions. Breath-hold duration was 13.5 seconds, and a respiratory sensor was used to monitor compliance. The entire protocol lasted 15-20 minutes.

Data for each 2-echo mDIXON sequence was reconstructed on the MR system using the available standard single-peak spectral model of fat. Each sequence yielded four images per slice: water only, fat only, in-phase, out-phase. From the DICOM “water” and “fat” images, a pixel-wise “signal fat-fraction” parameter map was generated as follows using an offline post-processing tool. Fat percentage was calculated as (SI (Fat) x 100)/(SI(W+F)) where SI is the pixel signal intensity on the reconstructed fat and water images. Images were independently analyzed by a trained abdominal radiologist, documenting direct reading of percentage of fat in all eight hepatic segments using an ROI of 200-300 mm. Care was taken to select only representative regions, avoiding hepatic vasculature, biliary structures, or regions with artefacts.

Results

Bioprofiles accurately predict MRI-assessed liver fat infiltration

Alignment of model predictions with PDFF was assessed across 3 metrics: AUROC at 5%, 15%, and Spearman coefficient. In the validation set, Bioprofiles outperformed foundation model (CLMBR[23,24]), clinical (FSI[15], HSI[16]), and technical benchmarks on all tasks (Table 2). Notably, CLMBR underperformed steatosis-specific clinical benchmarks at estimating PDFF.

To account for patterns of gaps in real-world data, we conducted sensitivity testing over data availability. We synthetically masked the fine-tuning cohort based on feature availability patterns learned from an external real-world EHR dataset to evaluate all models (Supplementary Table 2). Under these conditions, the Bioprofiles continued to outperform all benchmarks on all tasks.

To evaluate the generalizability of our models, we evaluated data from 5 clinical trials and hospital-based cohort studies where steatosis or MASLD status was assessed[29–33]. Bioprofiles continued to outperform clinical and technical benchmarks in all cases. (Table 3)

Bioprofiles facilitate efficient subject nomination in simulation

To understand the impact of model improvements on study logistics, we conducted study simulations on each of our validation sets. For most model/task combinations, Portfolio Selection tended to outperform standard sampling, and Bioprofiles outperformed clinical benchmarks. However, for the targeted recruitment task, Bioprofiles alone were sufficient to achieve strong performance, likely due to the simpler desired distribution (Supplementary Tables 3, 4). However, on the even recruitment task, we observed a synergistic effect between the use of Bioprofile patient representations and the Portfolio Selection algorithm. To identify 12 subjects with PDFF > 10%, the Bioprofile required 19 screened subjects compared to 39 (HSI) and 43 (FSI). To identify 27 subjects in the even recruitment task, Bioprofiles required only 30 subjects screened, compared to 60 subjects (HSI) and 69 subjects (FSI) (Table 4). Notably,

Bioprofile performance for the even recruitment task is close to ideal, with only three additional screenings projected to achieve recruitment goals. Study performance and variance for the even recruitment task was found to be a function of recruitment pool size (Supplementary Table 5).

Bioprofiles successfully nominated of subjects for a human study

Using our Portfolio Selection algorithm, we identified a recruitment strategy that would enable us to embed the targeted recruitment task within the even recruitment task with minimal projected impact on total recruitment numbers (Supplementary Figure 1C). Recruitment was therefore conducted in two waves: first, the cohort projected to have PDFF > 10% was recruited, followed by the remaining subjects needed to obtain even recruitment. We first computed Bioprofiles for 45,484 individuals who met inclusion criteria in the VUMC CDW, and over 3 months of outreach, contacted 1,388 individuals through the MHAV platform. The targeted recruitment task was achieved after 20 subjects were screened, and the even recruitment task after 9 additional subjects, for 29 total screens. These values are in line with study simulations (Table 4) and result in substantially reduced numbers of excess screens relative to existing standards (Figure 1A, B). Next, we examined model performance using continuous measures of liver fat abundance, independent of bin definitions, comparing the Bioprofile predicted PDFF distribution to i) expected background and ii) actual observed distributions (Figure 1C). The predicted distribution was found to be significantly different from the background (KS test: $p < 1E\text{-}3$) but not significantly different to the distribution observed in enrolled study participants (KS test: $p = 0.6$). Finally, Bioprofiles maintained strong subject level performance for PDFF among the recruited subjects (Figure 1D).

Discussion

Our study demonstrates that AI-driven patient profiling can meaningfully surface subclinical disease signatures hidden in routine clinical data, providing a basis for streamlining clinical trial recruitment and enabling more proactive patient care. Using Bioprofiles, we efficiently nominated candidate participants for a prospective pilot study at VUMC, achieving high yield and strong agreement with imaging findings. By leveraging routine health record data pre-emptively, our approach enabled targeted invitations to patients who were most likely to qualify, thereby reducing screening failures. Simulation studies showed that the Bioprofile nomination pipeline halved the number of subjects that had to be screened to enroll the same number of eligible participants compared with conventional risk scores or untargeted approaches.

The relatively poor performance of an EHR-based foundation model on the PDFF prediction task highlights one of the intrinsic limitations of real-world data sources. Because the features captured in the EHR are typically generated through expressions of clinical judgement[34], they do not provide complete representation of patient physiology. For diseases associated with substantial rates of underdiagnosis or clinical uncertainty, our approach of directly modeling accessible physiology is more likely to capture relevant disease states.

From a broader perspective, this work illustrates how a foundation model of cardiometabolic health can be translated into tangible recruitment benefits. While our study focused on trial recruitment, the underlying approach – estimating the results of less accessible tests using

more accessible features – has much wider potential applications. Most immediately, Bioprofile patient representations have potential for providing probabilistic estimates for a wide series of cardiometabolic endpoints. Bioprofile-generated probability distributions can be easily incorporated into health economic or net-benefit calculations with the goal of optimizing care delivery and resource allocation. More directly, by mapping the overlap between Bioprofile output and existing clinical practice guidelines, Bioprofiles could act as a source of risk stratification or decision support.

Our findings should be interpreted in light of certain limitations. Our external validation focused on populations from Asia and trial participants. We are currently planning more extended validation in American community health contexts. Our VUMC pilot included 29 screens, which, while sufficient to demonstrate feasibility, is a small sample to fully establish the model's performance characteristics. These subjects were also model-selected based on likelihood of meeting inclusion criteria and subject to response bias and so are not necessarily representative of the wider population. Our aggressive approach to feature mapping and curation was helpful in driving signal but was also highly labor intensive. Methods for interfacing or computing Bioprofiles from standard data models will be needed to enable wider-scale deployment. Finally, as currently configured, Bioprofiles are best suited to the highest prevalence indications among seemingly healthy primary care encounters. While we believe they can be extended to other indications, alternative datasets will be necessary to capture the full slate of "clinically accessible data" in other contexts.

In summary, this pilot highlights a promising future where AI-based physiological profiling becomes a routine tool for improving research efficiency and patient care. By leveraging routine health data to surface individuals with subclinical or unlabeled disease, Bioprofile modeling enabled precise and efficient recruitment in a MASLD-focused study, reducing the screening burden and improving alignment with inclusion criteria. This approach has the potential to accelerate drug development (through faster, more cost-effective trials) and to improve patient outcomes (through earlier detection and intervention). Beyond MASLD, applying Bioprofiles to identify subclinical cases could transform trial logistics and clinical care across a broad range of conditions by converting an invisible reservoir of patients into an actionable cohort for early identification, recruitment or discovery.

	Pretraining Cohort (1M)		**UKB Imaging Cohort (n=43k)**			**VUMC Cohort (Pre-Visit EHR Extract) (n=29) (2 ineligible for MRI)**			**VUMC Cohort (Study Visit) (n=29) (2 ineligible for MRI)**		
	Metric	Fraction with Data	PDFF < 5% (32k)	5% < PDFF < 15% (9.1k)	PDFF > 15% (2.1k)	PDFF < 5% (10)	5% < PDFF < 15% (9)	PDFF > 15% (9)	PDFF < 5% (11)	5% < PDFF < 15% (9)	PDFF > 15% (9)
Age (years)	49.61 (49.58 - 49.63)	1	55.01 (54.93 - 55.09)	55.47 (55.32 - 55.63)	54.1 (53.85 - 54.42)	50.06 (40.42 - 58.54)	41.45 (33.63 - 48.51)	43.28 (37.77 - 49.12)	50.55 (40.18 - 60)	41.49 (34.5 - 50.13)	43.55 (36.19 - 50.23)
Sex (fraction male)	0.486 (0.485 - 0.487)	1	0.439 (0.434 - 0.443)	0.606 (0.597 - 0.616)	0.576 (0.556 - 0.597)	0.618 (0.25 - 0.875)	0.533 (0.163 - 0.777)	0.644 (0.333 - 0.888)	0.618 (0.25 - 0.875)	0.533 (0.163 - 0.777)	0.644 (0.333 - 0.888)

BMI (kg/m2)	25.29 (25.28 - 25.3)	0.992	25.65 (25.61 - 25.69)	28.84 (28.76 - 28.93)	30.53 (30.36 - 30.77)	23.39 (21.57 - 26.34)	38.29 (33.91 - 43.88)	36.42 (32.78 - 40.98)	23.52 (21.12 - 26.31)	37.14 (32.41 - 41.44)	37.47 (33.05 - 40.95)
Triglycerides	1.55 (1.54 - 1.55)	0.955	1.47 (1.46 - 1.48)	2.04 (2.07 - 2.12)	2.29 (2.24 - 2.35)	0.683 (0.484 - 0.88)	1.48 (1.27 - 1.67)	2.67 (1.69 - 3.92)	0.698 (0.52 - 0.863)	1.47 (1.28 - 1.69)	2.94 (2.1 - 3.81)
Serum ALT (U/L)	23.11 (23.08 - 23.14)	0.923	20.7 (20.58 - 20.83)	28.12 (27.86 - 28.41)	35.68 (34.82 - 36.79)	21.69 (15.93 - 28.32)	49.16 (27.22 - 73.63)	71.81 (47.68 - 96.06)	21.66 (15.61 - 26.57)	48.37 (27.39 - 74.78)	78.8 (56.93 - 101.7)
Serum AST (U/L)	24.95 (24.93 - 24.98)	0.801	25 (24.91 - 25.13)	27.41 (27.16 - 27.67)	30.6 (30.1 - 31.16)	27.27 (21.18 - 34.45)	31.86 (24.94 - 42.19)	42.29 (30.15 - 54.65)	27.56 (21.93 - 33.32)	31.53 (25.05 - 39.51)	47.08 (33.88 - 64.11)
Diabetes Status	0.043 (0.043 - 0.044)	0.767	0.0168 (0.0154 - 0.0182)	0.0511 (0.0475 - 0.0561)	0.0704 (0.060 - 0.082)	0 (0 - 0)	0.57 (0.222 - 0.888)	0.755 (0.444 - 1)	0 (0 - 0)	0.539 (0.222 - 0.836)	0.793 (0.555 - 1)
PDFF	N/A	0	2.60 (2.59 - 2.61)	8.36 (8.31 - 8.41)	21.0 (20.7 - 21.2)	N/A	N/A	N/A	2.10 (1.73 - 2.46)	9.90 (7.59 - 11.7)	30.1 (24.9 - 35.0)

Table 1: Characteristics of the pretraining, fine-tuning, and recruited cohorts. All values are means (95% confidence interval) unless otherwise indicated.

	AUC (PDFF= 5%)	AUC (PDFF= 15%)	Spearman Correlation vs. PDFF
CLMBR (EHR-based Foundation Model)	0.693 (0.681-0.705)	0.716 (0.690-0.741)	0.361 (0.345-0.377)
Framingham Steatosis Index (FSI)	0.780 (0.762-0.798)	0.742 (0.732-0.753)	0.516 (0.502-0.530)
Hepatic Steatosis Index (HSI)	0.798 (0.792-0.803)	0.747 (0.742-0.752)	0.545 (0.534-0.556)
Gradient Boosted (BMI/ALT/AST)	0.819 (0.795-0.842)	0.752 (0.74-0.763)	0.554 (0.533-0.575)
Gradient Boosted (All Bioprofile input features)	0.817 (0.798-0.836)	0.761 (0.751-0.772)	0.576 (0.550-0.602)
Bioprofile	**0.849 (0.842-0.855)**	**0.801 (0.792-0.811)**	**0.646 (0.637-0.654)**

Table 2: Model performance for predicting PDFF. All values are means (95% confidence interval) unless otherwise indicated.

	Framingham Steatosis Index (FSI)	Hepatic Steatosis Index (HSI)	Gradient Boosted (All Bioprofile input features)	Bioprofile
FLINT[30] (Spearman R vs. LBX Steatosis) (n=283)	0.06 (-0.05-0.18)	0.12 (0-0.24)	0.11 (0-0.23)	**0.20 (0.07-0.32)**
Gifu Cohort[31] (AUROC vs ultrasound assessed MASLD status) (n=786)	0.86 (0.83-0.89)	0.84 (0.8-0.88)	0.85 (0.82-0.89)	**0.88 (0.85-0.91)**
Kaohsiung Cohort[32] (AUROC vs ultrasound assessed MASLD severity) (n=407)	0.69 (0.64-0.75)	0.68 (0.63-0.74)	0.68 (0.63-0.74)	**0.73 (0.68-0.78)**
Iksan Cohort[33] (Spearman R vs. CT-assessed disease severity) (n=913)	0.25 (0.18-0.31)	0.36 (0.29-0.42)	0.35 (0.29-0.41)	**0.41 (0.36-0.48)**
NAFLD Adult Database[29] (Spearman R vs. LBX Steatosis) (n=1266)	0.02 (-0.04-0.09)	0.06 (0-0.13)	0.14 (0.07-0.20)	**0.23 (0.17-0.3)**
Dandelion Cohort (AUROC vs. MRI-assessed presence of steatosis) (n=176)	0.596 (0.535-0.659)	0.676 (0.618-0.733)	0.744 (0.672-0.816)	**0.761 (0.691-0.831)**

Table 3: External model validation on fatty liver status/biopsy (LBX)-confirmed steatosis. All values are means (95% confidence interval) unless otherwise indicated.

	Targeted Recruitment (12 with PDFF > 10%)		Even Recruitment (9 each with i) 0% ≤ PDFF < 5%, ii) 5% ≤ PDFF 15%, iii) 15% ≤ PDFF)	
Random Sampling	116 (98-143)		161 (127-217)	
	Standard Sampling	Portfolio Selection	Standard Sampling	Portfolio Selection
FSI	43 (40-44)	37 (36-38)	69 (67-71)	50 (49-51)
HSI	39 (36-41)	34 (33-35)	60 (58-62)	47 (46-47)
Bioprofile	**19 (16-22)**	**19 (19-20)**	44 (43-45)	**30 (30-30)**

Table 4: Simulated Recruitment Requirements for the fine-tuning set. All values are means (95% confidence interval) unless otherwise indicated.

Figure 1: Results of VUMC Bioprofile Deployment Pilot: A, B) Bar charts of included/excluded subjects for targeted (A) and even (B) recruitment. Naïve sampling, clinical score and Bioprofile projected represent simulations from Table 4. C) Kernel density estimates for the background PDFF distribution (blue), Bioprofile predicted PDFF distribution (orange), and observed PDFF distribution (green). D) Subject level model performance on VUMC-recruited cohorts, stratified by input data type.

Bibliography

1. A multisociety Delphi consensus statement on new fatty liver disease nomenclature - PubMed. Accessed May 13, 2025. https://pubmed.ncbi.nlm.nih.gov/37363821/

2. Nielsen EM, Anderson KP, Marsden J, Zhang J, Schreiner AD. Nonalcoholic Fatty Liver Disease Underdiagnosis in Primary Care: What Are We Missing? *J Gen Intern Med*. 2022;37(10):2587-2590. doi:10.1007/s11606-021-07197-3

3. Younossi ZM, Koenig AB, Abdelatif D, Fazel Y, Henry L, Wymer M. Global epidemiology of nonalcoholic fatty liver disease-Meta-analytic assessment of prevalence, incidence, and outcomes. *Hepatol Baltim Md*. 2016;64(1):73-84. doi:10.1002/hep.28431

4. American Association of Clinical Endocrinology Clinical Practice Guideline for the Diagnosis and Management of Nonalcoholic Fatty Liver Disease in Primary Care and Endocrinology Clinical Settings: Co-Sponsored by the American Association for the Study of Liver Diseases (AASLD) - PubMed. Accessed May 13, 2025. https://pubmed.ncbi.nlm.nih.gov/35569886/

5. Alexander M, Loomis AK, Fairburn-Beech J, et al. Real-world data reveal a diagnostic gap in non-alcoholic fatty liver disease. *BMC Med*. 2018;16(1):130. doi:10.1186/s12916-018-1103-x

6. Chalasani N, Younossi Z, Lavine JE, et al. The diagnosis and management of nonalcoholic fatty liver disease: Practice guidance from the American Association for the Study of Liver Diseases. *Hepatol Baltim Md*. 2018;67(1):328-357. doi:10.1002/hep.29367

7. Castera L, Friedrich-Rust M, Loomba R. Noninvasive Assessment of Liver Disease in Patients With Nonalcoholic Fatty Liver Disease. *Gastroenterology*. 2019;156(5):1264-1281.e4. doi:10.1053/j.gastro.2018.12.036

8. Harrison SA, Allen AM, Dubourg J, Noureddin M, Alkhouri N. Challenges and opportunities in NASH drug development. *Nat Med*. 2023;29(3):562-573. doi:10.1038/s41591-023-02242-6

9. Stine JG, Munaganuru N, Barnard A, et al. Change in MRI-PDFF and Histologic Response in Patients With Nonalcoholic Steatohepatitis: A Systematic Review and Meta-Analysis. *Clin Gastroenterol Hepatol*. 2021;19(11):2274-2283.e5. doi:10.1016/j.cgh.2020.08.061

10. Flint A, Andersen G, Hockings P, et al. Randomised clinical trial: semaglutide versus placebo reduced liver steatosis but not liver stiffness in subjects with non-alcoholic fatty liver disease assessed by magnetic resonance imaging. *Aliment Pharmacol Ther*. 2021;54(9):1150-1161. doi:10.1111/apt.16608

11. Resmetirom (MGL-3196) for the treatment of non-alcoholic steatohepatitis: a multicentre, randomised, double-blind, placebo-controlled, phase 2 trial - The Lancet. Accessed May 13, 2025. https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(19)32517-6/abstract?utm_source=chatgpt.com

12. Idilman IS, Aniktar H, Idilman R, et al. Hepatic Steatosis: Quantification by Proton Density Fat Fraction with MR Imaging versus Liver Biopsy. *Radiology*. 2013;267(3):767-775. doi:10.1148/radiol.13121360

13. Tamaki N, Ajmera V, Loomba R. Non-invasive methods for imaging hepatic steatosis and their clinical importance in NAFLD. *Nat Rev Endocrinol*. 2022;18(1):55-66. doi:10.1038/s41574-021-00584-0

14. Davison BA, Harrison SA, Cotter G, et al. Suboptimal reliability of liver biopsy evaluation has implications for randomized clinical trials. *J Hepatol*. 2020;73(6):1322-1332. doi:10.1016/j.jhep.2020.06.025

15. Long MT, Pedley A, Colantonio LD, et al. Development and Validation of the Framingham Steatosis Index to Identify Persons with Hepatic Steatosis. *Clin Gastroenterol Hepatol Off Clin Pract J Am Gastroenterol Assoc*. 2016;14(8):1172-1180.e2. doi:10.1016/j.cgh.2016.03.034

16. Lee JH, Kim D, Kim HJ, et al. Hepatic steatosis index: a simple screening tool reflecting nonalcoholic fatty liver disease. *Dig Liver Dis Off J Ital Soc Gastroenterol Ital Assoc Study Liver*. 2010;42(7):503-508. doi:10.1016/j.dld.2009.08.002

17. Schattenberg JM, Balp MM, Reinhart B, et al. NASHmap: clinical utility of a machine learning model to identify patients at risk of NASH in real-world settings. *Sci Rep*. 2023;13(1):5573. doi:10.1038/s41598-023-32551-2

18. He Y, Huang F, Jiang X, et al. Foundation Model for Advancing Healthcare: Challenges, Opportunities and Future Directions. *IEEE Rev Biomed Eng*. 2025;18:172-191. doi:10.1109/RBME.2024.3496744

19. Novo Nordisk. Wegovy Insert ID 5342984. https://www.accessdata.fda.gov/drugsatfda_docs/label/2024/215256s011lbl.pdf

20. UK Biobank - UK Biobank. Accessed May 13, 2025. https://www.ukbiobank.ac.uk/

21. CDC. National Health and Nutrition Examination Survey. National Health and Nutrition Examination Survey. April 30, 2025. Accessed May 13, 2025. https://www.cdc.gov/nchs/nhanes/index.html

22. Starekova J, Hernando D, Pickhardt PJ, Reeder SB. Quantification of Liver Fat Content with CT and MRI: State of the Art. *Radiology*. 2021;301(2):250-262. doi:10.1148/radiol.2021204288

23. Steinberg E, Jung K, Fries JA, Corbin CK, Pfohl SR, Shah NH. Language models are an effective representation learning technique for electronic health record data. *J Biomed Inform*. 2021;113:103637. doi:10.1016/j.jbi.2020.103637

24. Wornow M, Thapa R, Steinberg E, Fries JA, Shah NH. EHRSHOT: An EHR Benchmark for Few-Shot Evaluation of Foundation Models. Published online December 11, 2023. doi:10.48550/arXiv.2307.02028

25. Caussy C, Alquiraish MH, Nguyen P, et al. Optimal threshold of controlled attenuation parameter with MRI-PDFF as the gold standard for the detection of hepatic steatosis. *Hepatol Baltim Md*. 2018;67(4):1348-1359. doi:10.1002/hep.29639

26. Romero-Gómez M, Lawitz E, Shankar RR, et al. A phase IIa active-comparator-controlled study to evaluate the efficacy and safety of efinopegdutide in patients with non-alcoholic fatty liver disease. *J Hepatol*. 2023;79(4):888-897. doi:10.1016/j.jhep.2023.05.013

27. Cunha GM, Thai TT, Hamilton G, et al. Accuracy of common proton density fat fraction thresholds for magnitude- and complex-based chemical shift-encoded MRI for assessing hepatic steatosis in patients with obesity. *Abdom Radiol N Y*. 2020;45(3):661-671. doi:10.1007/s00261-019-02350-3

28. Yuan W, Beaulieu-Jones BK, Yu KH, et al. Temporal bias in case-control design: preventing reliable predictions of the future. *Nat Commun*. 2021;12:1107. doi:10.1038/s41467-021-21390-2

29. Tonascia J. Nonalcoholic Fatty Liver Disease Adult Database 2. Published online December 3, 2024:8635907 B. doi:10.58020/EQRT-BX74

30. Tonascia J. The Farnesoid X Receptor Ligand Obeticholic Acid in NASH Treatment. Published online January 13, 2023:10206968 MB. doi:10.58020/H3M1-Z998

31. Fukuda T, Hamaguchi M, Kojima T, et al. Association between serum γ-glutamyltranspeptidase and atherosclerosis: a population-based cross-sectional study. *BMJ Open*. 2014;4(10):e005413. doi:10.1136/bmjopen-2014-005413

32. Hsieh CJ, Wang PW, Hu TH. Association of Adiponectin Gene Polymorphism with Nonalcoholic Fatty Liver Disease in Taiwanese Patients with Type 2 Diabetes. *PLoS ONE*. 2015;10(6):e0127521. doi:10.1371/journal.pone.0127521

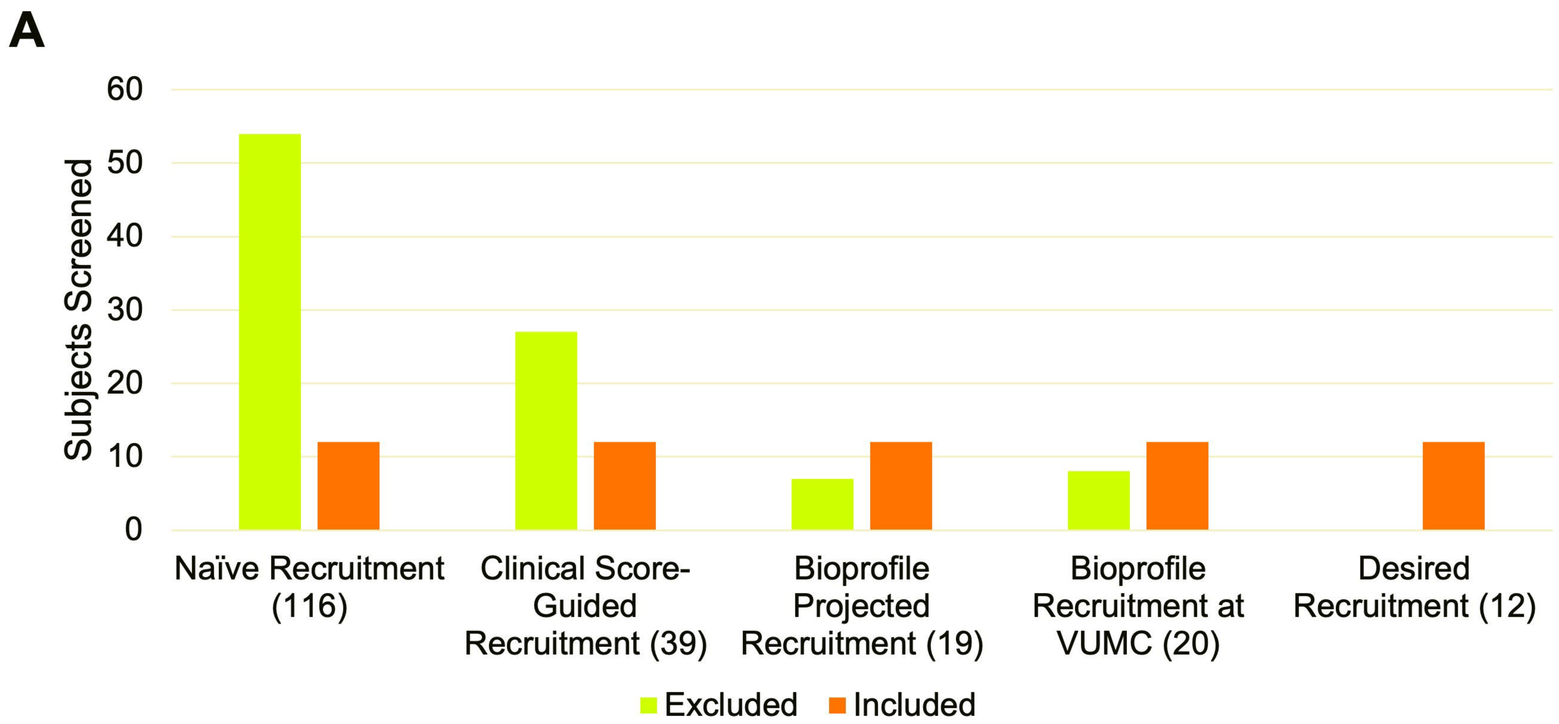
33. Han AL. Association of Cardiovascular Risk Factors and Metabolic Syndrome with non-alcoholic and alcoholic fatty liver disease: a retrospective analysis. *BMC Endocr Disord*. 2021;21(1):91. doi:10.1186/s12902-021-00758-x

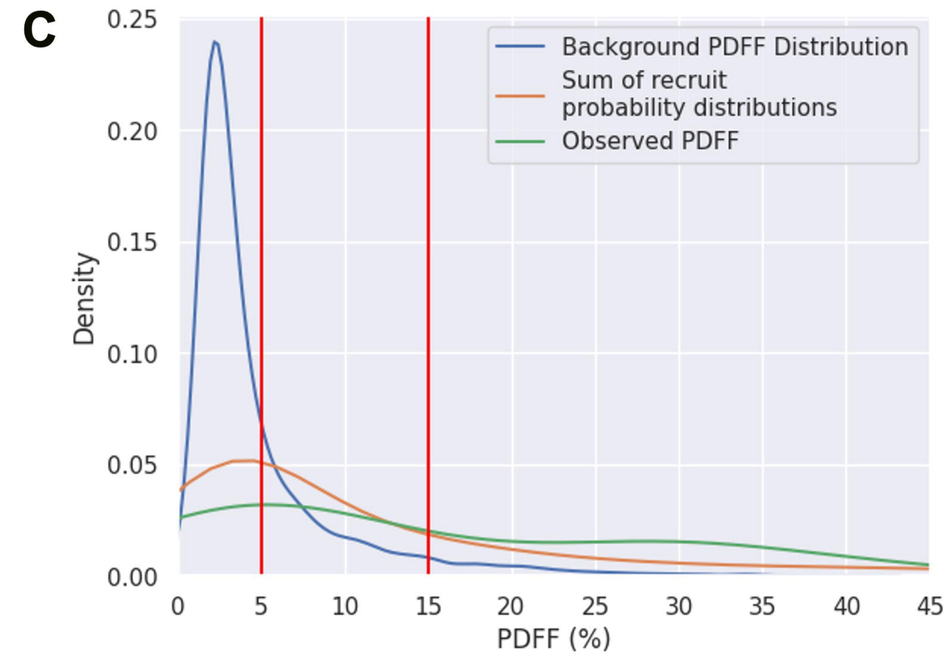
34. Beaulieu-Jones BK, Yuan W, Brat GA, et al. Machine learning for patient risk stratification: standing on, or looking over, the shoulders of clinicians? *NPJ Digit Med*. 2021;4:62. doi:10.1038/s41746-021-00426-3

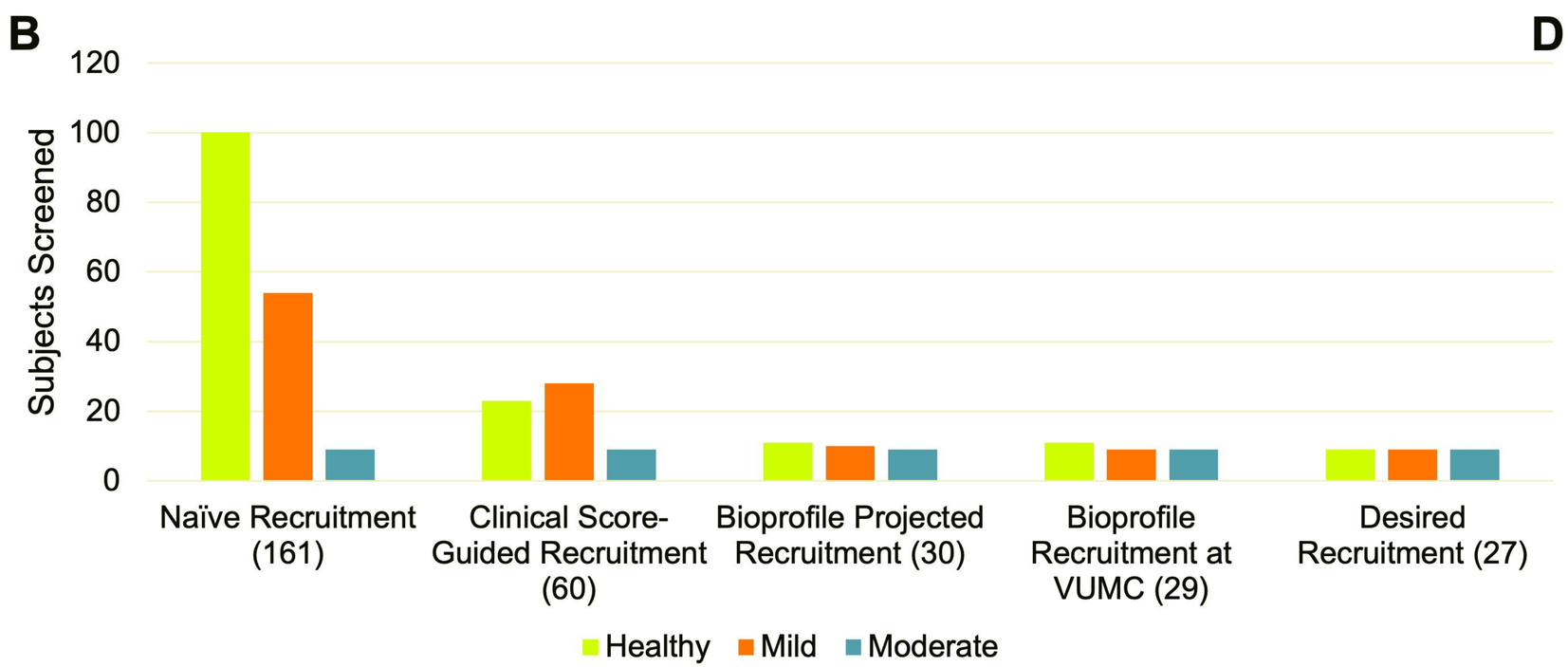
Acknowledgements

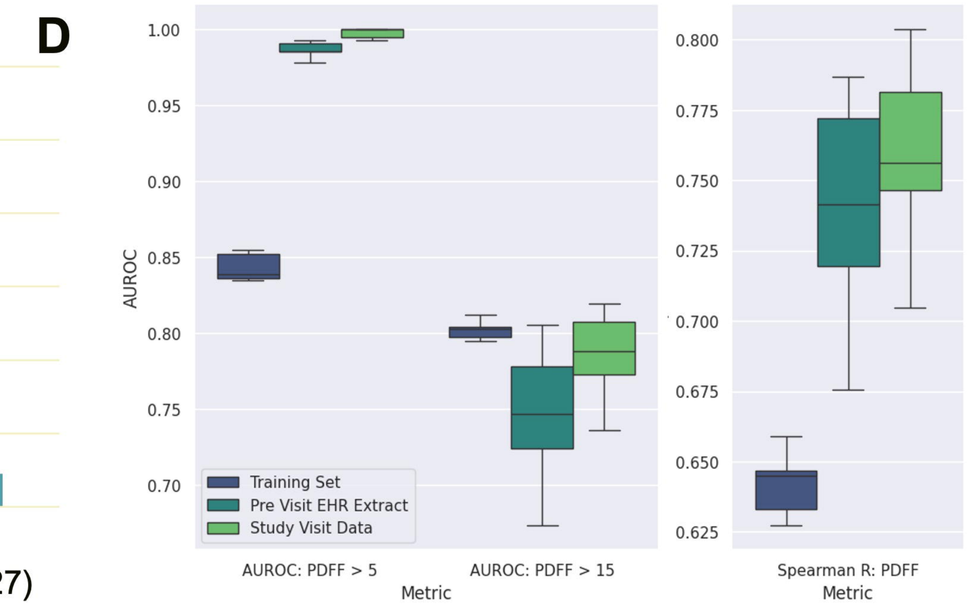
This research has been conducted using data from UK Biobank, a major biomedical database under application ID (76125).

The FLINT and NAFLD Adult Database studies was conducted by the study investigators and supported by the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK). The resources from the FLINT (https://doi.org/10.58020/h3m1-z998) and NAFLD Adult Database (https://doi.org/10.58020/eqrt-bx74) studies reported here were supplied by NIDDK Central Repository (NIDDK-CR) and are available for request at https://repository.niddk.nih.gov. This manuscript was not prepared under the auspices of the FLINT and NAFLD Adult Database studies and does not necessarily reflect the opinions or views of the FLINT and NAFLD Adult Database studies, NIDDK-CR, or NIDDK.


A
Subjects Screened
60
50
40
30
20
10
0
Naïve Recruitment (116)
Clinical Score-Guided Recruitment (39)
Bioprofile Projected Recruitment (19)
Bioprofile Recruitment at VUMC (20)
Desired Recruitment (12)
Excluded
Included


C
Density
0.25
0.20
0.15
0.10
0.05
0.00
0
5
10
15
20
25
30
35
40
45
PDFF (%)
Background PDFF Distribution
Sum of recruit probability distributions
Observed PDFF


B
Subjects Screened
120
100
80
60
40
20
0
Naïve Recruitment (161)
Clinical Score-Guided Recruitment (60)
Bioprofile Projected Recruitment (30)
Bioprofile Recruitment at VUMC (29)
Desired Recruitment (27)
Healthy
Mild
Moderate


D
AUROC
1.00
0.95
0.90
0.85
0.80
0.75
0.70
Training Set
Pre Visit EHR Extract
Study Visit Data
AUROC: PDFF > 5
AUROC: PDFF > 15
Metric
0.800
0.775
0.750
0.725
0.700
0.675
0.650
0.625
Spearman R: PDFF
Metric