The Journal of Clinical Endocrinology & Metabolism, 2024, **109**, e2090–e2099
https://doi.org/10.1210/clinem/dgae037
Advance access publication 23 January 2024
Clinical Research Article

The Lipoprotein Profile Evaluated by 1H-NMR Improves the Performance of Genetic Testing in Familial Hypercholesterolemia

Daiana Ibarretxe,[1,2] Dídac Llop,[1,2] Cèlia Rodríguez-Borjabad,[1,2] Natalia Andreychuk,[1,2] Núria Plana,[1,2] Roberto Scicali,[3] Ana González-Lleó,[1,2] Núria Amigó,[4,5,2] Josefa Girona,[1,2] and Lluís Masana[1,2]

[1]Unitat Medicina Vascular I Metabolisme, Unitat de Recerca en Lípids i Arteriosclerosi, Hospital Universitari Sant Joan, Universitat Rovira i Virgili, IISPV, 43201 Reus, Spain
[2]Centro de Investigación Biomédica en Red de Diabetes y Enfermedades Metabólicas Asociadas (CIBERDEM), Instituto de Salud Carlos III (ISCIII), 28029 Madrid, Spain
[3]Department of Clinical and Experimental Medicine, Internal Medicine, Garibaldi-Nesima Hospital, University of Catania, 95122 Catania, Italy
[4]Biosfer Teslab, 43201 Reus, Spain
[5]Department of Basic Medical Sciences, Institut d'Investigació Sanitària Pere Virgili (IISPV), Universitat Rovira i Virgili (URV), 43204 Reus, Spain

Correspondence: Dídac Llop Paredes, PhD, Facultat de Medicina i Ciències de la Salut, Universitat Rovira i Virgili, Carrer Sant Llorenç, 21, 43204 Reus, Spain. Email: didac.llop@urv.cat.

Abstract

Background: The familial hypercholesterolemia (FH) diagnosis is based on clinical and genetic criteria. A relevant proportion of FH patients fulfilling the criteria for definite FH have negative genetic testing. Increasing the identification of true genetic-based FH is a clinical challenge. Deepening the analysis of lipoprotein alterations could help increase the yield of genetic testing. We evaluated whether the number, size, and composition of lipoproteins assessed by 1H-NMR could increase the identification of FH patients with pathogenic gene variants.

Methods: We studied 294 clinically definite FH patients, 222 (75.5%) with positive genetic testing, as the discovery cohort. As an external validation cohort, we studied 88 children with FH, 72 (81%) with positive genetic testing. The advanced lipoprotein test based on 1H-NMR (Liposcale®) was performed at baseline after a lipid-lowering drug washout of at least 6 weeks. The association of variables with genetic variants was evaluated by random forest and logistic regression. Areas under the curve (AUCs) were calculated. A predictive formula was developed and applied to the validation cohort.

Results: A formula derived from nuclear magnetic resonance (NMR) lipoprotein analyses improved the identification of genetically positive FH patients beyond low-density lipoprotein (LDL)-cholesterol levels (AUC = 0.87). The parameters contributing the most to the identification formula were LDL particle number, high-density lipoprotein size, and remnant cholesterol. The formula also increases the classification of FH children with a pathogenic genetic variation.

Conclusion: NMR lipoprotein profile analysis identifies differences beyond standard lipid parameters that help identify FH with a positive pathogenic gene variant, increasing the yield of genetic testing in FH patients.

Key Words: familiar hypercholesterolemia, genetic testing, nuclear magnetic resonance, machine learning, cardiovascular risk

Familial hypercholesterolemia (FH) is a genetic disorder that results in elevated low-density lipoprotein cholesterol (LDL-C) levels from birth and the subsequent progression of atherosclerotic cardiovascular (CV) disease (1, 2). Among the pathogenic FH-associated genetic variants responsible for FH, the most common ones affect the LDL receptor gene (LDLR), which accounts for approximately 85% to 90% of cases. Pathogenic gene variants in other genes, such as apolipoprotein B (APOB) and proprotein convertase subtilisin/kexin type 9 (PCSK9), are less common causes of FH (3). Recent epidemiological data suggest a FH prevalence of 1 in 200 to 250 people (4). However, the exact percentage of genetic diagnosis in FH depends on the population studied and the diagnostic criteria used. A definite clinical FH diagnosis can be established by meeting the Dutch Lipid Clinic Network (DLCN) criteria or similar criteria based on LDL-C levels, personal and family history, and physical signs (5, 6). However, a pathogenic FH gene variant cannot be detected in all definite clinical FH diagnoses. In general, it is estimated that FH-associated genetic variants are detected in 70% to 90% of FH cases (3). The remaining cases within the spectrum of FH syndrome (7) are a mixture of polygenic alterations, unknown genetic factors, and multifactorial causes driven by environmental-genetic interactions (8). Importantly, the *vera* FH diagnosis requires genetic testing. Beyond identifying the specific genetic variant responsible for the disorder, positive

Received: 19 October 2023. **Editorial Decision:** 14 January 2024. **Corrected and Typeset:** 1 February 2024

genetic testing helps identify family members who may be at risk of developing FH and can inform personalized treatment and management plans (9), thereby increasing patient empowerment and adherence to therapy (10). Patients with genetically confirmed FH show a higher prevalence of CV events than nongenetically confirmed patients with similar LDL-C concentrations (11-13). FH patients carrying a pathogenic variant have a 22-fold increased risk for coronary artery disease, while clinical FH without a pathogenic variant increases the risk by 6% despite having similar LDL concentrations (14). Moreover, patients with genetically confirmed FH show a higher prevalence of coronary artery calcification than patients without genetically confirmed FH (15). Therefore, the presence of genetic alterations indicates a group of FH patients at higher risk of requiring a more intensive clinical approach (16). The identification of FH patients with positive genetic testing should be a priority for CV risk stratification. However, the cost of genetic testing is high and not universally reimbursed by public health systems (4). The relatively low performance of genetic studies even in patients with clinically definite FH is a clinical and economical challenge, and several studies have provided additional clues to refine the clinical characteristics of those clinical FH patients carrying a pathogenic gene variant (17, 18). We hypothesized that genetically positive and negative clinically diagnosed FH could have small metabolic differences undetectable by standard biochemical lipid profiles but detectable by metabolomic/lipidomic techniques that could improve the identification of genetically positive patients, improving the yield of genetic testing and helping to stratify CV risk.

In this study, we applied 2D 1-H nuclear magnetic resonance (NMR) spectroscopy to serum samples (Liposcale® test) to study the advanced lipoprotein profile of FH patients with and without positive genetic testing results (19).

Our aim was to assess whether the advanced lipoprotein profile assessed by NMR can properly classify genetically positive and negative FH patients and provide an equation that could be used in clinical settings to classify new patients.

Methods

Study Design and Patients

We performed a baseline cross-sectional study of 2 cohorts.

Cohort 1 (adult FH)

This cohort included 294 patients attending our Lipid Unit due to heterozygous familial hypercholesterolemia with a score in the DLCN equal to or greater than 6 (75.6% with FH-associated pathogenic genetic variants). Anamnesis, anthropometry, and physical examination evaluating corneal arch and Achillean tendon xanthomas and blood tests, including serum 1H-NMR analysis, were obtained at baseline. Patients taking statins or other cholesterol-lowering agents underwent a 6-week washout period to avoid the modifications in the advanced lipoprotein profile induced by lipid-lowering agents.

Cohort 2 (children FH)

This cohort included 88 children and adolescents aged 4 to 18 years attending our Lipid Unit due to heterozygous FH. They were participating in the Early Detection of Familial Hypercholesterolemia in Children (DECOPIN Project, ClinicalTrials.gov Identifier: NCT04370899), which focused on the implementation of opportunistic, direct, and reverse cascade FH screening. Children were diagnosed with FH if they had a positive genetic test (82%) or LDL-C > 150 mg/dL and 1 of the parents had definite FH. At the time of inclusion, no patients were receiving lipid-lowering therapy. An exhaustive medical history, including family CV and dyslipidaemia history, a complete physical examination, and anthropometry data, was recorded. To calculate body mass index (BMI) in children, we used the BMI score, which was calculated by the following equation: [(BMI children - BMI 50th percentile of Orbegozo's growth curves)/SD 50th percentile of Orbegozo's growth curves].

An overnight fasting blood sample was obtained from each patient to perform the biochemical studies.

The study fulfilled the criteria of the Helsinki Declaration, and the Ethical and Clinical Investigation Committee of the Pere Virgili Institute for Health Research approved the study. All participants signed the written consent form.

Standard Biochemical Variables

Plasma and serum aliquots were stored at −80 °C in our research institute's (Pere Virgili Institute for Health Research) biobank until the measurements. Biochemical parameters, lipids, and apolipoproteins were measured using colorimetric, enzymatic assays (Spinreact, SA, Spain; Wako Chemicals GmbH, Germany), which were adapted to the Cobas Mira Plus Autoanalyser (Roche Diagnostics, Spain).

Genetic Testing

In both cohorts, the presence of FH-associated variants was studied by next-generation sequencing (Liponext, SEQPROLIPO RS, Roche Diagnostics). Liponext detects variants in LDLR, APOB, PCSK9, LIPA, APOE, STAP1 (ADH), and LDLRAP1 (ARH) genes and copy-number variation in LDLR.

2D-1H-NMR Lipoprotein and Glycoprotein Profile Evaluation

A 200-ml aliquot of plasma was used for lipoprotein analysis by 2D-1H-NMR (Liposcale®) (19). The particle size and number of 9 subtypes of lipoproteins [large, medium, and small very low density lipoprotein (VLDL), LDL, and high-density lipoprotein (HDL)] as well as cholesterol and triglyceride concentrations in lipoprotein subclasses were determined as previously reported. Remnant cholesterol (Rem-Chol) was calculated as the sum of intermediate-density lipoprotein, and VLDL cholesterol. Briefly, particle concentrations were obtained from the measured amplitudes and attenuation of their spectroscopically distinct methyl group signals using the 2D diffusion ordered NMR spectroscopy pulse. By this method, the hydrodynamic characteristics of the molecules associated with each subclass of lipoprotein can be measured. From the diffusion coefficients, the sizes of different subclasses of lipoproteins are directly calculated through the Stokes–Einstein equation. The direct measurement of the size, as in this method, is of particular importance since it is used to calculate the number of lipoprotein particles. The methyl signal was surface fitted with 9 Lorentzian functions associated with each lipoprotein subtype. The area of each Lorentzian function reflected the lipid concentration of

each lipoprotein subtype, and the size of each subtype was calculated from the diffusion coefficient. The particle number of each lipoprotein subtype was calculated by dividing the lipid volume by the particle volume of a given class. Lipid volumes were determined using common conversion factors to convert concentration units into volume units. The variation coefficients for particle number were between 2% and 4%. The variation coefficients for particle size were lower than 0.3%.

Statistical Methods

Data procedure and prediction models

Cohort 1 was used to develop the different prediction models to assess whether the Liposcale®2D-1H-NMR lipid profile or the conventionally measured LDL-C, along with age, sex, and BMI, could differentiate between genetically positive and negative patients. Additionally, it was also used for internal validation. Cohort 2 was used as the external validation cohort. All numeric variables were standardized. Due to the small number of missing values (3 out of 294 for BMI in cohort 1), they were replaced by the mean of all available values for the same variable in the dataset. The cohort 1 dataset was randomly divided into a training dataset, validation dataset, and test dataset in a 0.7/0.15/0.15 ratio. The training dataset was used to train the classification models, tune the hyperparameters of the algorithms, and discard unimportant features. The validation dataset was used to choose the best classifying model. The test dataset and cohort 2 were used as internal and external validation datasets, respectively. The test set and the external validation set were always independent of the training and tuning process during the development of the models. The prediction algorithms used were random forest (RF) models and logistic regression (LR) models.

RF

RF is a bagging ensemble classification algorithm based on the growth of multiple conditional inference trees. Feature selection was performed using the *Boruta* package, which is one of the wrappers built around the RF algorithm to obtain the importance of the variables based on the mean decrease in accuracy (20).

LR

LR models predict the probability of a binary dependent variable using the maximum likelihood estimation to determine regression coefficients. As several variables were highly correlated and LR is highly affected by multicollinearity, we performed elastic net regressions to select the most influential features (21).

Since class imbalance existed between patients who were genetically positive and negative (4:1), all adjusted models were weighted to increase the importance of the minority class. The optimal hyperparameters were those that maximized the area under the curve (AUC) of the different receiver operating characteristic curves (ROCs) in the models for each method, which were chosen using 10-fold cross-validation.

Performance metrics of internal and external validations

The evaluation of the performance of the different models in the internal test set and external set was studied with the AUC of the ROC curves, together with the accuracy, sensitivity, specificity, positive predictive value, negative predictive value, and F_1 score. We also provide the confusion matrix of each prediction model.

Model distributions

To avoid opportunistic splits of the dataset that could lead to unrealistic results, we performed model distributions with each predictive algorithm fit. We arbitrarily took 100 different seeds that created 100 different splits from the internal dataset. We calculated 100 different AUCs of the internal and external test sets of each obtained training model. With the 100 AUCs obtained, we calculated a density distribution from which we obtained an average AUC and its respective confidence interval (CI). The results obtained from the predictions were considered valid if they were within the 95% CIs obtained from the model distributions. All AUCs for the internal and external test sets are given with their respective 95% CIs derived from the model distributions.

Descriptive and statistical analyses

Continuous normal variables are presented as the mean and SD, while continuous nonnormal variables are presented as the median and interquartile range. Categorical variables are presented as the percentage and number of individuals. T tests of unpaired data and Mann-Whitney tests were used to compare normal and nonnormal variables, respectively, in cohorts 1 and 2. To compare normal variables between the training, validation, and test, we used one-way ANOVA in the response variables that were normally distributed or the Kruskal-Wallis test if the response variable was nonnormally distributed. Categorical variables were compared using the chi-square test. P-values < .05 were considered to indicate statistical significance. To control the overall type I error rate across all comparisons, Bonferroni P-value correction was applied when multiple comparisons were performed between patients with a detected genetic variant and those without a genetic variant in cohorts 1 and 2. The corrected P-value was obtained by multiplying the standard P-value by the number of comparisons performed (40 in each cohort). Statistical analyses were performed in R Studio, version 4.0.1.

Results

Characteristics of the Cohorts

Table 1 shows the general characteristics of cohort 1 sorted by genetic data. Cohort 1 included 294 hypercholesterolaemic patients, 222 of whom were genetically positive (75.5%) for FH. Their mean age was 40 (32-49), 53% of them were male, and their average BMI was 25.64 (23-28). Patients with a genetic positive test were younger and had higher LDL-C and more LDL particles. The LDL and HDL sizes were also larger. Genetically negative patients show more triglycerides and more total, large, medium, and small VLDL particles. They also had more small HDL particles. They had more VLDL and remnant cholesterol. VLDL and HDL were triglyceride enriched. When Bonferroni corrections were applied, we observed that most of the differences regarding the lipoprotein profile were maintained.

Supplementary Table S1 (22) shows the general characteristics of the training set, the internal validation set, and the test set. The clinical and biochemical characteristics of

Table 1. Description of cohort 1

	Overall cohort 1 (n = 294)	Gen. positive test (n = 222)	Gen. negative test (n = 72)	*P*-value	Bonferroni corrected *P*-value
Age (years)	40 (32-49)	37 (30-49)	44 (37.75-52)	<0.0001	.006
Sex, male (%)	156 (53)	51.58	58.57	0.44	1
BMI (kg/m^2)	25.64 (23-28)	25.37 (22.69-37.40)	26.67 (23.92-28.71)	0.02	.8
DLCN	9 (7-12)	9 (6-12)	9.50 (8.75-11.25)	0.02	.8
Smoke, n (%)	103 (35.03)	80 (36.03)	23 (31.94)	0.62	1
DMT2, n (%)	3 (1.02)	3 (1.35)	0 (0)	0.75	1
Hypertension, n (%)	30 (10.20)	19 (8.56)	11 (15.28)	0.16	1
SBP (mmHg)	120 (110-130)	120 (110-130)	122 (115-138)	0.06	1
DBP (mmHg)	75 (70-80)	74 (70-80)	80 (70-84)	0.02	.8
Previous CV disease, n, (%)	8 (2.72)	5 (2.25)	3 (4.17)	0.66	1
Xanthomas, n (%)	113 (38.43)	76 (34.23)	37 (51.39)	0.01	.4
Corneal arch, n (%)	108 (36.73)	67 (30.18)	41 (56.94)	<0.0001	.003
Total cholesterol (mg/dL)	317.1 (281.1-363.5)	336.4 (286.9 -375.1)	294.4 (265.8-322.9)	<0.0001	.001
LDL-C (CM) (mg/dL)	243 (207.5-294)	261 (217-312.4)	211 (187-239.9)	<0.0001	<.0001
HDL-C (CM) (mg/dL)	50.27 (42.25-60.32)	49.11 (39.54-59.45)	54.07 (46.85-62.11)	0.0001	.04
TG (mg/dL)	97 (70.86-141.72)	88.57 (62-126.66)	125.50 (88.57-164.41)	<0.0001	<.0001
Lipoprotein particle number					
VLDL (nmol/L)	31.31 (20.54-56.16)	28.69 (17.30-48.86)	48.60 (30.20-74.32)	<0.0001	<.0001
Large VLDL (nmol/L)	0.64 (0.44-1.16)	0.58 (0.37-0.94)	1.02 (0.62-1.50)	<0.0001	<.0001
Medium VLDL (nmol/L)	3.55 (2.39-5.76)	3.25 (2.14-4.99)	5.59 (3.39-7.15)	<0.0001	<.0001
Small VLDL (nmol/L)	27.54 (17.79-48.98)	24.96 (14.97-42.91)	42.15 (26.55-64.96)	<0.001	<.0001
LDL (nmol/L)	2037 (1803-2339)	2074 (1851-2436)	1917 (1739-2178)	<0.0001	.03
Large LDL (nmol/L)	310 (270.5-354.6)	321.4 (281.3-372.7)	277.4 (245.5-304.3)	<0.0001	<.0001
Medium LDL (nmol/L)	795.5 (655.4-982.5)	822.4 (692.1-1024.5)	685 (582.1-840.3)	<0.0001	<.0001
Small LDL (nmol/L)	925.2 (827.7-1083.7)	926.9 (818.3-1086.3)	912.4 (856.4-1062.8)	0.99	1
HDL (μmol/L)	26.66 (23.38-30.45)	25.77 (22.93-29.71)	28.77 (25.95-31.15)	<.0001	.005
Large HDL (μmol/L)	0.37 (0.33-0.42)	0.37 (0.33 -0.42)	0.35 (0.32-0.40)	0.02	.8
Medium HDL (μmol/L)	10.16 (9.11-11.21)	10.33 (9.29-11.49)	9.47 (8.90-10.82)	0.008	.32
Small HDL (μmol/L)	15.99 (13.19-19.23)	15.17 (12.69-18.30)	18.77 (16.48-28.53)	<0.0001	<.0001
Lipoprotein composition (mg/dL)					
VLDL-C	9.47 (4.33-20.13)	8.72 (3.95-17.44)	17.92 (8.51-25.33)	<0.0001	<.0001
VLDL-TG	42.76 (29.07-75.08)	38.88 (24.46-60.32)	65.98 (42.50-102.30)	<0.0001	<.0001
IDL-C	13.49 (10.11-18.34)	12.96 (9.19-17.94)	14.99 (12.18-19.75)	0.003	.12
IDL-TG	11.79 (8.53-15.04)	10.74 (7.89-14.13)	13.99 (11.66-17.29)	<0.0001	<.0001
Remnant-C	23.87 (16.63-36.74)	21.26 (14.62-34.51)	31.15 (22.76 -45.05)	<0.0001	.0002
LDL-C	209.3 (186-244.6)	216.3 (192.1-254.6)	193.1 (172.9-217.6)	<0.0001	<.0001
LDL-TG	25.92 (21.30-31.51)	26.15 (21.30-31.92)	24.78 (21.27-29.84)	0.2	1
HDL-C	57.67 (51.39-64.94)	58.10 (51.24-65.78)	56.44 (52.34-61.86)	0.39	1
HDL-TG	8.99 (5.20-13.35)	7.86 (4.43-12.64)	11.62 (9.01-16.60)	<0.0001	<.0001
Lipoprotein size (diameter, nm)					
VLDL	42.08 (41.98-42.22)	42.09 (41.99-42.22)	42.05 (41.94-42.18)	0.07	1
LDL	21.43 (21.21-21.57)	21.47 (21.30-21.58)	21.23 (21.01-21.42)	<0.0001	<.0001
HDL	8.33 (8.27-8.38)	8.35 (8.29-8.39)	8.26 (8.21-8.31)	<0.0001	<.0001

Abbreviations: BMI, body mass index; C, cholesterol; CM, conventionally measured; CV, cardiovascular; DBP, diastolic blood pressure; DLCN, Duch Lipid Clinic Network (the scores shown do not include points for positive genetic testing; not applicable for children); DMT2, diabetes mellitus type 2; HDL, high-density lipoprotein; IDL, immediate-density lipoprotein; LDL, low-density lipoprotein; nmol/L, nanomolar; nm, nanometer; SBP, systolic blood pressure; TG, triglycerides; VLDL, very low density lipoprotein; μmol/L, micromolar.
Normally distributed data is expressed in terms of mean and SD. Nonnormally distributed data is expressed in terms of median and interquartile range. Categorical data is expressed in percentages. To evaluate differences between groups, t-tests were used to compare normally distributed data, the Mann–Whitney U test was used to compare nonnormally distributed data, and the chi-square (χ^2) test was used to compare categorical data. *P*-values < .05 are considered statistically significant.

Table 2. Description of cohort 2

	Overall cohort 2 (n = 88)	Gen. positive (n = 72)	Gen. negative (n = 16)	*P*-value	Bonferroni corrected *P*-value
Age (years)	9 (6-12.25)	9 (6-13)	8.50 (7-10)	0.75	1
Sex, male (%)	47 (53.40)	52.74	34.78	.01	0.4
BMI score (kg/m^2)	18.86 (16.07-21.74)	19.30 (16.34-21.79)	16.44 (15.54-20.63)	.17	1
Smoke, n (%)	4 (4.54)	4 (5.55)	0 (0)	.76	1
SBP (mmHg)	108 (99.5-117)	109 (100-117)	103 (97.25-111)	.18	1
DBP (mmHg)	63 (57.50-70)	63 (58-70)	63.50 (55.75-68.50)	.57	1
Xanthomas, n (%)	8 (9.09)	7 (9.72)	1 (6.25)	1	1
Corneal arch, n (%)	1 (1.14)	1 (1.39)	0 (0)	1	1
Total cholesterol (mg/dL)	262 (228.8-295.2)	266.5 (229.8-305.5)	234 (227-254.5)	.04	1
LDL-C (CM) (mg/dL)	186 (149.8-219.2)	195.5 (159.8-229)	153 (145.8-164.8)	<.0001	0.14
HDL-C (CM) (mg/dL)	59 (50.75-67)	56 (50-66)	65.50 (59.75-76.25)	.04	1
Triglycerides (mg/dL)	64 (50-88)	60 (47.75-87.25)	79.50 (53.75-91.25)	.17	1
Lipoprotein particle number					
VLDL (nmol/L)	24.70 (18.20-38.72)	23.55 (17.25-37.75)	33.81 (23.73-45.73)	.04	1
Large VLDL (nmol/L)	0.75 (0.51-0.99)	0.71 (0.49-0.88)	0.92 (0.73-1.12)	.02	0.8
Medium VLDL (nmol/L)	3.30 (2.32-4.24)	3.20 (2.19-4.10)	3.95 (2.86-4.96)	.04	1
Small VLDL (nmol/L)	20.54 (15.24-33.29)	19.81 (14.62-32.20)	29.31 (19.90-38.77)	.053	1
LDL (nmol/L)	1806 (1481.3-2081.2)	1851.3 (1515.2 -2155.3)	1581 (1442-1758)	.053	1
Large LDL (nmol/L)	260.8 (215.2-304.5)	272.4 (224.6-318.9)	225 (206.7-256.4)	.01	0.4
Medium LDL (nmol/L)	615.6 (464-775.7)	663.4 (469.7-824.1)	518 (432.5-602.1)	.03	1
Small LDL (nmol/L)	850.6 (776.2-1010.6)	878.3 (777.3-1017.7)	800 (767.8-877.1)	.17	1
HDL (µmol/L)	27 (23.91-30.35)	26.95 (23.75-29.70)	28.73 (25.37-32.08)	.09	1
Large HDL (µmol/L)	0.28 (0.25-0.31)	0.28 (0.25-0.31)	0.28 (0.25-0.30)	.77	1
Medium HDL (µmol/L)	9.41 (8.62-10.23)	9.37 (8.62-10.24)	9.52 (8.49-10.10)	.95	1
Small HDL (µmol/L)	17.43 (15.13-19.75)	17.17 (14.93-19.09)	19.44 (17.04-21.28)	.03	1
Lipoprotein composition (mg/dL)					
VLDL-C	9.95 (6.35-13.21)	9.32 (5.66-13.02)	11.67 (9.94-17.13)	.03	1
VLDL-TG	33.80 (23.87-52.15)	31.18 (23.63-48.38)	43.48 (30.25-59.60)	.049	1
IDL-C	7.78 (5.69-9.91)	7.58 (5.59-9.73)	8.44 (6.89-10.45)	.27	1
IDL-TG	7.63 (5.85-9.94)	7.40 (5.70-8.87)	8.05 (6.90-11.15)	.1	1
Remnant-C	17.70 (12.82-23.98)	16.39 (12.59-22.86)	20.57 (16.11-27.01)	.05	1
LDL-C	184.56 (152.94-215.23)	194.74 (162.22-227.90)	160.2 (151-178.1)	.02	0.80
LDL-TG	18.18 (14.16-24.02)	18.57 (14.01-24.78)	15.80 (14.48-21.37)	.34	1
HDL-C	57.24 (48.93-64.66)	57.24 (48.93-64.56)	57.35 (50.80-59.37)	.69	1
HDL-TG	8.08 (5.64-11.39)	7.10 (5.10-10.01)	12.59 (9.40-17.88)	<.0001	0.009
Lipoprotein size (diameter, nm)					
VLDL	42.31 (42.01-42.50)	42.32 (42.01-42.51)	42.29 (42.05-42.45)	.73	1
LDL	21.33 (21.15-21.52)	21.34 (21.14-21.57)	21.32 (21.20-21.44)	.38	1
HDL	8.30 (8.25-8.33)	8.31 (8.26-8.34)	8.26 (8.24-8.29)	.005	0.2

Abbreviations: BMI, body mass index; C, cholesterol; CM, conventionally measured; DBP, diastolic blood pressure; DLCN, Duch Lipid Clinic Network (the scores shown do not include points for positive genetic testing; not applicable for children); HDL, high-density lipoprotein; IDL, immediate-density lipoprotein; LDL, low-density lipoprotein; nmol/L, nanomolar; nm, nanometer; SBP, systolic blood pressure; TG, triglycerides; VLDL, very low density lipoprotein; µmol/L, micromolar.
Normally distributed data is expressed in terms of mean and SD. Nonnormally distributed data is expressed in terms of median and interquartile range. Categorical data is expressed in percentages. To evaluate differences between groups, t-tests were used to compare normally distributed data, the Mann–Whitney U test was used to compare nonnormally distributed data, and the chi-square (χ^2) test was used to compare categorical data. *P*-values < .05 are considered statistically significant.

patients distributed in these 3 datasets were similar, except for LDL diameter ($P = .01$).

The proportion of genetically positive FH patients was 74.4% for the training set, 71.42% for the validation set, and 85.1% for the test set. No differences were found between the 3 datasets ($P = .23$).

Table 2 shows the characteristics of cohort 2, which included 88 hypercholesterolaemic children, of whom 72 were genetically positive (82%). Their mean age was 9 (6-12.25), 53.40% were male, and their average BMI score was 18.86 (16.07-21.74). Patients with a genetically positive test result had more LDL-C and LDL large, medium, and small particles.

Genetically negative patients had more VLDL particles, more VLDL and remnant cholesterol, and smaller HDL particles. When Bonferroni corrections were applied, most of the significative differences were lost.

Hyperparameterization, Feature Selection and Packages Used

The optimal hyperparameters tuned for each model and method are shown in Supplementary Table S2 (22). Supplementary Table S2 (22) also shows the most important features selected for RF and LR with their importance ranking.

Internal Validation

Table 3 shows the confusion matrix and performance metrics obtained from the predictions of the internal test dataset with the 1H-NMR profile as well as the AUC of the ROC with its 95% CI obtained from the model distribution. The predictions obtained from the 1H-NMR profile with RF and LR show similar performance metrics with good predictive power [RF: AUC = 0.88 (95% CI: .73–.96); LR: AUC = 0.87 (95% CI: .69–.96)]. Both models show high sensitivity and specificity (RF: sensitivity = 0.86 and specificity = 0.86; LR: sensitivity = 0.86 and specificity = 0.69), with slightly increased specificity in RF. In the same direction, the RF model also shows a better F_1 score than LR (RF: F_1 score = 0.67; LR: F_1 score = 0.51). LDL-C conventionally measured predictions also showed good results in the internal dataset [RF: AUC 0.86 (95% CI: .66–.93); LR: AUC (95% CI: .83 [0.71-0.93] (Supplementary Table S3 (22)]. The ROC curves of the internal test predictions calculated with the 1H-NMR profile are shown in Fig. 1A. ROC curves of the internal predictions calculated with conventionally measured LDL-C are shown in Supplementary Fig. S1A (22).

External Validation

Table 3 shows the confusion matrix and performance metrics obtained of the predictions of the external dataset with the 1H-NMR profile with their AUCs and their respective 95% CI obtained from the model distribution. The predictions obtained from the 1H-NMR profile with RF and LR also show similar results, with good predictive power for both methods [RF: AUC = 0.83 (95% CI: .72–.84); LR: AUC = 0.82 (95% CI: .82–.84)]. The RF model showed slightly decreased sensitivity compared with LR (RF: sensitivity = 0.63; LR: sensitivity = 0.75) but increased specificity (RF: specificity = 0.83; LR: specificity = 0.72). Regarding the F_1 score, both models show very similar results (RF: F_1 score = 0.53; LR: F_1 score = 0.51). However, conventionally measured LDL showed noticeably worse results [RF: AUC = 0.60 (95% CI: .56-.72); LR: AUC = 0.59 (95% CI: .57–.61) (Supplementary Table S3 (22)]. The ROC curves of the external dataset predictions calculated with the 1H-NMR profile are shown in Fig. 1B. ROC curves of the external predictions calculated with conventionally measured LDL-C are shown in Supplementary Fig. S1B (22).

Clinical Application

Based on logistic regression prediction, we formulated an algorithm with potential clinical utility that could help distinguish between genetically positive and negative FH patients. The result of the algorithm is a probability, which, if it is

Table 3. Confusion matrix and performance metrics of the adjusted models with the Liposcale variables, age, sex

Model	Confusion matrix		AUC (95 CI%)	Sensitivity (95% CI)	Specificity (95% CI)	PPV (95% CI)	NPV (95% CI)	Accuracy (95% CI)	F1 score
Internal validation									
RF	–	– 6	0.88 (0.73-0.96)	0.86 (0.70-0.95)	0.86 (0.42-1)	0.55 (0.23-0.83)	0.97 (0.83-1)	0.86 (0.71-0.95)	0.67
	+	+ 1							
		+ 30							
LR	–	– 6	0.87 (0.69-0.96)	0.86 (0.42-1)	0.69 (0.51-0.83)	0.36 (0.14-0.62)	0.96 (0.80-1)	0.74 (0.58-0.86)	0.51
	+	+ 1							
		+ 24							
External validation									
RF	–	– 10	0.83 (0.72-0.84)	0.63 (0.35-0.85)	0.83 (0.73-0.91)	0.45 (0.24-0.68)	0.91 (0.81-0.97)	0.80 (0.70-0.87)	0.53
	+	+ 6							
		+ 60							
LR	–	– 12	0.82 (0.82-0.84)	0.75 (0.48-0.93)	0.72 (0.60-0.82)	0.38 (0.23-0.59)	0.93 (0.83-0.98)	0.73 (0.62-0.82)	0.51
	+	+ 4							
		+ 52							

Abbreviations: –, negative genetic test; +, positive genetic test; AUC, area under the curve; CI, confidence interval; LR, logistic regression; NPV, negative predictive value; PPV, positive predictive value; RF, random forest.

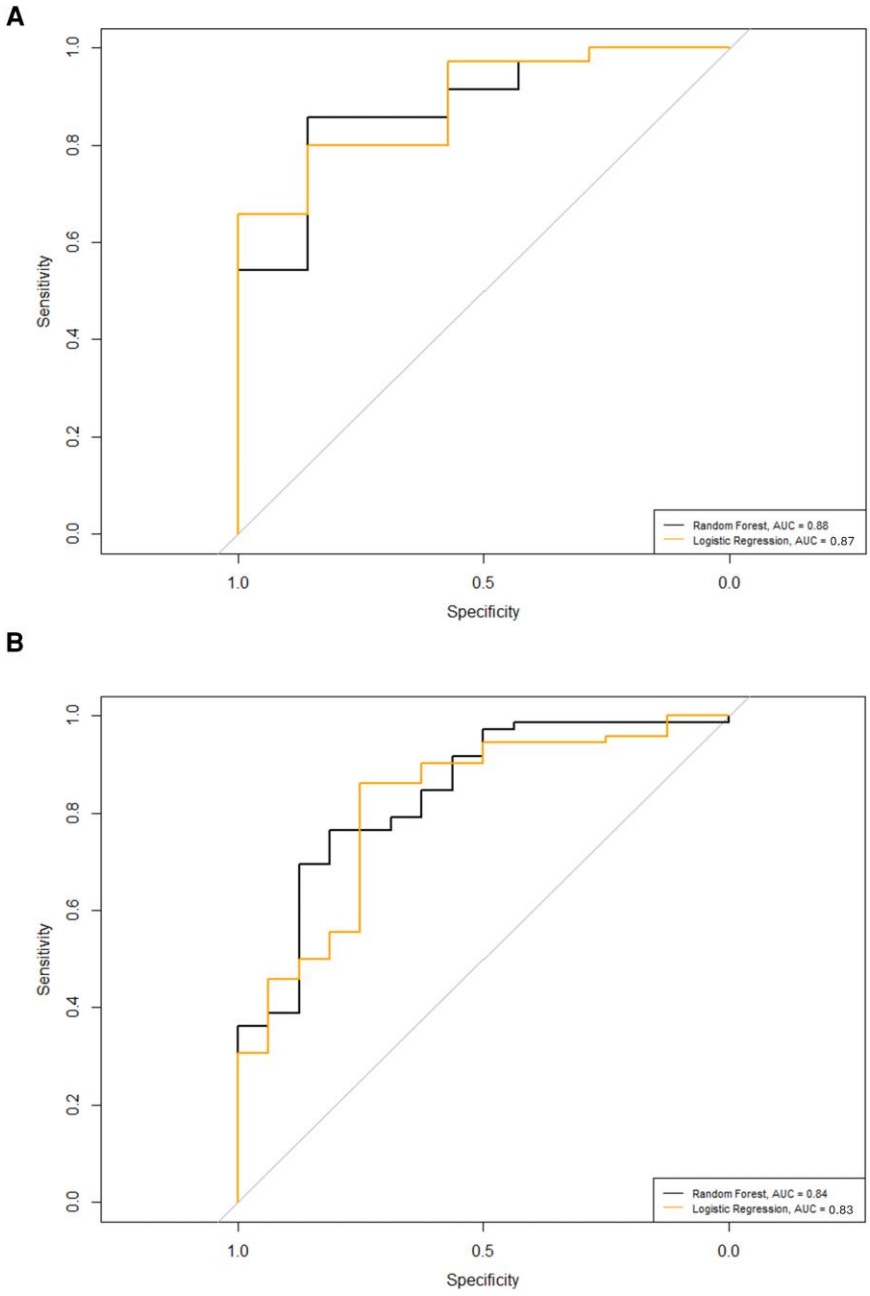

Figure 1. Receiver operating characteristic curves for the predictions of the internal dataset (A) and external dataset (B) with the Liposcale variables.

greater than 0.5, will cause the patient to present a genetically positive test. Otherwise, it will be considered negative:

$$p = \frac{1}{1 + e^{-(0.38 + 1.02 \cdot HDLz - 0.70 \cdot Rem\text{-}Chol + 0.61 \cdot LargeLDL)}}$$

where *HDLz* represents the mean diameter of the HDL particles, *Rem-Chol* represents the remnant cholesterol (intermediate-density lipoprotein + VLDL-C), and *LargeLDL* represents the number of large LDL particles.

Discussion

Our work addresses a challenging clinical issue: improving the performance of genetic testing to diagnose true monogenic FH patients. A study performed in our region showed that only 41.1% of patients with DLCN scores of > 6 carry a pathogenic variant, and this percentage increased up to 48.5% in patients with a DLCN ≥ 8. On the other hand, in FH-suspicious patients with a DLCN < 6, 80% had no genetic variant, but 20% of these patients who were supposed to be at a lower FH probability were genetically positive (23). Moreover, a recent publication from Denmark almost replicates these data, showing that among clinically definite FH patients who underwent genetic testing, an FH-associated genetic variant was identified in approximately 40% (24). According to similar studies in those patients with definite FH, a pathogenic variant is identified in 60% to 80%; in possible FH, the performance is even lower (21% to 44%) (25-28). In highly suspicious-FH children, the yield of genetic testing ranges from 60% to 95% (29, 30).

There are several reasons for this mismatch. One is the inaccuracy of clinical indices, including DLCN, Simon Broome, or MedPed, with relatively low sensitivity and specificity (31). Moreover, the implementation of indices is usually difficult. Data on family history are usually difficult to gather, and some physical signs, such as corneal arch, are not meaningful at the usual age of disease detection, or lesions, such as xanthomas, are not present because of the generalized use of statins for high, although undiagnosed, cholesterol concentrations (32). An example is the Spanish FH SAFEHEART cohort, where xanthomas were present in <15% of cases and the corneal arch was present in approximately 30% of the patients (33). Moreover, the standard clinical indices cannot be applied to children because they have lower LDL-C levels, and, in general, there are neither physical signs yet nor personal or family history of CV events (34).

The important gap between monogenic FH suspicion and the real genetic disease leads to several clinical inconveniences. It is difficult to explain to the patient and family the absence of a clear genetic cause. The need for family screening is different, although the treatment, driven by LDL-C levels, is often quite intensive (35). In addition, evidence has shown that FH patients with genetically positive tests have a 22-fold increased risk of coronary artery disease compared to the general population, while the increase in risk of genetically unconfirmed patients is 6%. On the other hand, patients with genetically confirmed FH show more coronary artery calcification, which is a strong marker of CV risk (12, 15).

In some countries, the FH diagnosis is also the gate to reimbursement for specialized therapies such as PCSK9 inhibitors (36), and the lack of genetic confirmation generates some confusion in patients, physicians, health authorities, and payers (37, 38). To overlook this point, we suggested including all clinically or genetically diagnosed FH patients under the concept of FH syndrome (7). Moreover, performing genetic tests with a negative rate of 50% is unsustainable from an economical point of view.

In this study, we tested whether a deeper look at changes in the lipid profile by 1H-NMR taking into account not only cholesterol concentrations but also lipoprotein particle number, size, and composition could give us a tool to improve the performance of genetic testing. Interestingly, patients classified as probable or definite FH with negative genetic testing had in general more triglyceride-rich lipoprotein particles and more small HDL, suggesting that, although the markers were within the normal ranges, a component of atherogenic dyslipidaemia could explain a bias in diagnosis. After all the classification studies, we developed a formula based on those parameters that were more closely associated with the presence of genetic variants beyond the LDL-C concentrations, particularly in patients with LDL-C levels that are not as high and less disease-defining in themselves. After a comprehensive data analysis including clinical, biochemical, and NMR data, the parameters that were finally included in the formula to help classify positive and negative genetic FH patients were large LDL particle number, HDL size, and Rem-Chol obtained by NMR, which were probably acting as statistical surrogates of other lipoprotein parameters. According to the equation, a value above 0.5 classifies the patient as positive for genetic testing. For values above 0.5, the closer they were to 1, the more likely it was that the genetic test would be positive. In our study, according to AUC values (Table 2), the formula correctly classified 87% of our cohort of adult FH patients.

For external validation, we applied the formula to a cohort of children with FH. Interestingly, although we could expect different metabolic determinants in children compared to adults, the formula also classified up to 82% of these children correctly, validating our results and extending the use of the formula to younger ages.

The application of 1H-NMR to lipoprotein studies in FH could improve the classification of real monogenic FH patients including those with gene variants in *LDL-R, APOB,* and *PCSK9* genes (7-9). The 1H-NMR lipoprotein test and algorithm should be recommended in the identification of patients with LDL-C levels less disease-defining. Taking into account that the NMR study costs approximately one-quarter of genetic testing, this diagnostic step could also be highly efficient.

Our study has some limitations: the sample size is limited, particularly the child cohort; however, it was large enough to obtain statistically significant results. Moreover, our results should be validated in more independent cohorts. The lack of standardization in the NMR methods to measure lipoprotein particles limits the application of these results to the use of the Liposcale test, which is available for clinical use across Europe (www.biosferteslab.com) (39). As strengths, we would like to note the robustness of the analytical method and the results obtained.

The advanced lipid profile assessed by 1H-NMR detects lipid variations not observed by standard measurements that could help identify real monogenic FH, increasing the yield of genetic testing. Performing this test before indicating a genetic test will increase the performance of genetic studies and will save money.

Acknowledgments

We acknowledge the patients involved in the study for their essential collaboration. Moreover, we acknowledge the IISPV-BIOBANC for sample handling, storage, and custody.

Funding

This study was funded in part by a grant from the Fundació la Marató de TV3 on cardiovascular diseases (grant number 20152430).

Disclosure

D.I. received fees for lectures from Sanofi, Amryt, Sobi. N.A. has a patent ("Method for Lipoprotein Characterization") issued and is cofounder and stoke owner of Biosfer Teslab, the company analyzing the lipoprotein profile reported in the manuscript. L.M. received fees for lectures and/or advisory work from Amarin, Amryt, Daiichi-Sankyo, Ferrer, MSD, Novartis, Sanofi, Ultragenix, Viatrix, and is cofounder and stoke owner of Biosfer Teslab. D.L., C.R.-B., N.A., A.G.-L., N.P., R.S. declare no known conflicts of interest regarding the present work.

Data Availability

Data are available for interested researchers under previous petition on the corresponding author.

References

1. Cuchel M, Bruckert E, Ginsberg HN, *et al.* Homozygous familial hypercholesterolaemia: new insights and guidance for clinicians to improve detection and clinical management. A position paper fromthe consensus panel on familial hypercholesterolaemia of the European atherosclerosis society. *Eur Heart J*. 2014;35(32):2146-2157.
2. Mcgowan MP, Hamed S, Dehkordi H, Duell PB. Diagnosis and treatment of heterozygous familial. *J Am Heart Assoc*. 2019;8:e013225.
3. Watts GF, Gidding SS, Hegele RA, *et al.* International atherosclerosis society guidance for implementing best practice in the care of familial hypercholesterolaemia. *Nat Rev Cardiol*. 2023;20:845-869.
4. Tokgozoglu L, Kayikcioglu M. Familial hypercholesterolemia: global burden and approaches. *Curr Cardiol Rep*. 2021;23(10):151.
5. Sturm AC, Knowles JW, Gidding SS, *et al.* Clinical genetic testing for familial hypercholesterolemia. *J Am Coll Cardiol*. 2018;72(6):662-680.
6. Defesche JC, Lansberg PJ, Umans-Eckenhausen MA, Kastelein JJ. Advanced method for the identification of patients with inherited hypercholesterolemia. *Semin Vasc Med*. 2004;4(01):59-65.
7. Masana L, Ibarretxe D, Rodríguez-Borjabad C, *et al.* Toward a new clinical classification of patients with familial hypercholesterolemia: one perspective from Spain. *Atherosclerosis*. 2019;287:89-92.
8. Di Taranto MD, Giacobbe C, Palma D, *et al.* Genetic spectrum of familial hypercholesterolemia and correlations with clinical expression: implications for diagnosis improvement. *Clin Genet*. 2021;100(5):529-541.
9. Santos RD, Gidding SS, Hegele RA, *et al.* Defining severe familial hypercholesterolaemia and the implications for clinical management: a consensus statement from the international atherosclerosis society severe familial hypercholesterolemia panel. *Lancet Diabetes Endocrinol*. 2016;4(10):850-861.
10. Mostaza JM, Pintó X, Armario P, *et al.* SEA 2022 standards for global control of cardiovascular risk. *Clin Investig Arterioscler*. 2022;34(3):130-179.
11. Guijarro C, Civeira F, Masana L. Genetic confirmation of monogenic familial hypercholesterolemia advises a more intensive lipid-lowering approach. *JAMA Cardiol*. 2020;5(12):1452-1453.
12. Khan TZ, Breen J, Neves E, Grocott-Mason R, Barbir M. Prevalence of cardiovascular events in genetically confirmed versus unconfirmed familial hypercholesterolaemia. *Glob Cardiol Sci Pract*. 2020;2020(2):e202024.
13. Ahmad Z, Li X, Wosik J, *et al.* Premature coronary heart disease and autosomal dominant hypercholesterolemia: increased risk in women with LDLR mutations. *J Clin Lipidol*. 2016;10(1):101-108.e3.
14. Khera AV, Won H-H, Peloso GM, *et al.* Diagnostic yield and clinical utility of sequencing familial hypercholesterolemia genes in patients with severe hypercholesterolemia. *J Am Coll Cardiol*. 2016;67(22):2578-2589.
15. Descamps OS, Gilbeau JP, Luwaert R, Heller FR. Impact of genetic defects on coronary atherosclerosis in patients suspected of having familial hypercholesterolaemia. *Eur J Clin Invest*. 2003;33(1):1-9.
16. Mosley JD, Gupta DK, Tan J, *et al.* Predictive accuracy of a polygenic risk score compared with a clinical risk score for incident coronary heart disease. *JAMA*. 2020;323(7):627-635.
17. Lázaro P, Pérez de Isla L, Watts GF, *et al.* Cost-effectiveness of a cascade screening program for the early detection of familial hypercholesterolemia. *J Clin Lipidol*. 2017;11(1):260-271.
18. Marusic T, Sustar U, Sadiq F, *et al.* Genetic and clinical characteristics of peatients with homozygous and compound heterozygous familial hypercholesterolemia from three different populations: case series. *Front Genet*. 2020;11:572176.
19. Mallol R, Amigó N, Rodríguez MA, *et al.* Liposcale: a novel advanced lipoprotein test based on 2D diffusion-ordered 1H NMR spectroscopy. *J Lipid Res*. 2015;56(3):737-746.
20. Petkovic D, Altman R, Wong M, Vigil A. Improving the explainability of random forest classifier—user centered approach. *Pac Symp Biocomput*. 2018;23:204-215.
21. Engebretsen S, Bohlin J. Statistical predictions with glmnet. *Clin Epigenetics*. 2019;11(1):123.
22. Dídac L. Supplementary material "The lipoprotein profile evaluated by 1H-NMR improves the performance of genetic testing in familial hypercholesterolemia". *J Clin Endocrinol Metab*. 2023. https://github.com/DidacLlop/Supplementary-Data-JCEM
23. Martín-Campos JM, Plana N, Figueras R, *et al.* Autosomal dominant hypercholesterolemia in catalonia: correspondence between clinical-biochemical and genetic diagnostics in 967 patients studied in a multicenter clinical setting. *J Clin Lipidol*. 2018;12(6):1452-1462.
24. Hedegaard BS, Bork CS, Kanstrup HL, *et al.* Genetic testing increases the likelihood of a diagnosis of familial hypercholesterolaemia among people referred to lipid clinics: danish national study. *Atherosclerosis*. 2023;373:10-16.
25. Haralambos K, Whatley SD, Edwards R, *et al.* Clinical experience of scoring criteria for familial hypercholesterolaemia (FH) genetic testing in Wales. *Atherosclerosis*. 2015;240(1):190-196.
26. Silva PRS, Jannes CE, Oliveira TGM, *et al.* Evaluation of clinical and laboratory parameters used in the identification of index cases for genetic screening of familial hypercholesterolemia in Brazil. *Atherosclerosis*. 2017;263:257-262.
27. Amor-Salamanca A, Castillo S, Gonzalez-Vioque E, *et al.* Genetically confirmed familial hypercholesterolemia in patients with acute coronary syndrome. *J Am Coll Cardiol*. 2017;70(14):1732-1740.
28. Taylor A, Wang D, Patel K, *et al.* Mutation detection rate and spectrum in familial hypercholesterolaemia patients in the UK pilot cascade project. *Clin Genet*. 2010;77(6):572-580.
29. Van Der Graaf A, Avis HJ, Kusters DM, *et al.* Molecular basis of autosomal dominant hypercholesterolemia: assessment in a large cohort of hypercholesterolemic children. *Circulation*. 2011;123(11):1167-1173.
30. Klančar G, Grošelj U, Kovač J, *et al.* Universal screening for familial hypercholesterolemia in children. *J Am Coll Cardiol*. 2015;66(11):1250-1257.
31. Wilemon KA, Patel J, Aguilar-Salinas C, *et al.* Reducing the clinical and public health burden of familial hypercholesterolemia: a global call to action. *JAMA Cardiol*. 2020;5(2):217-229.
32. Kindt I, Mata P, Knowles JW. The role of registries and genetic databases in familial hypercholesterolemia. *Curr Opin Lipidol*. 2017;28(2):152-160.
33. Perez De Isla L, Alonso R, Watts GF, *et al.* Attainment of LDL-cholesterol treatment goals in patients with familial hypercholesterolemia: 5-year SAFEHEART registry follow-up. *J Am Coll Cardiol*. 2016;67(11):1278-1285.
34. Ibarretxe D, Rodríguez-Borjabad C, Feliu A, Bilbao JÁ, Masana L, Plana N. Detecting familial hypercholesterolemia earlier in life by actively searching for affected children:the DECOPIN project. *Atherosclerosis*. 2018;278:210-216.
35. Knickelbine T, Lui M, Garberich R, Miedema MD, Strauss C, VanWormer JJ. Familial hypercholesterolemia in a large ambulatory population: statin use, optimal treatment, and identification for advanced medical therapies. *J Clin Lipidol*. 2016;10(5):1182-1187.
36. Fogacci F, Giovannini M, Grandi E, *et al.* Management of high-risk hypercholesterolemic patients and PCSK9 inhibitors reimbursement policies: data from a cohort of Italian hypercholesterolemic outpatients. *J Clin Med*. 2022;11(16):4701.
37. Medeiros AM, Bourbon M. Genetic testing in familial hypercholesterolemia: is it for everyone? . *Curr Atheroscler Rep*. 2023;25(4):127-132.

38. Jenkins N, Lawton J, Douglas M, *et al*. How do index patients participating in genetic screening programmes for familial hypercholesterolemia (FH) interpret their DNA results? A UK-based qualitative interview study. *Patient Educ Couns*. 2013;90(3): 372-377.

39. Pintó X, Masana L, Civeira F, *et al*. Consensus document of an expert group from the spanish society of arteriosclerosis (SEA) on the clinical use of nuclear magnetic resonance to assess lipoprotein metabolism (liposcale®). *Clin Investig Arterioscler*. 2020;32(5): 219-229.