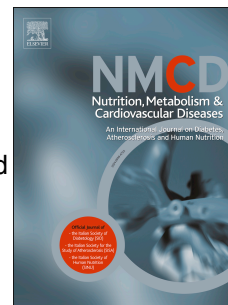# Journal Pre-proof

The Role of Adiposity, Diet and Inflammation on the Discordance between LDL-C and Apolipoprotein B

Richard J. Webb, Mohsen Mazidi, Gregory Y.H. Lip, Andre P. Kengne, Maciej Banach, Ian G. Davies
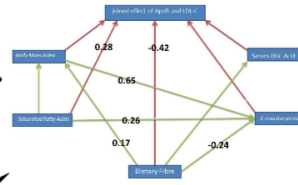
Please cite this article as: Webb RJ, Mazidi M, Lip GYH, Kengne AP, Banach M, Davies IG, The Role of Adiposity, Diet and Inflammation on the Discordance between LDL-C and Apolipoprotein B, *Nutrition, Metabolism and Cardiovascular Diseases*, https://doi.org/10.1016/j.numecd.2021.12.004.

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

1. Data categorised by ApoB / LDL-C discordance

High ApoB / High LDL-C | Low ApoB / High LDL-C
CONCORDANT | DISCORDANT
DISCORDANT | CONCORDANT
High ApoB / Low LDL-C | Low ApoB / Low LDL-C

■ Triglyceride ■ Cholesterol

2. Machine learning and SEMs applied to data

3. Mendelian randomisation applied to data

4. Adiposity found to be causal of ApoB
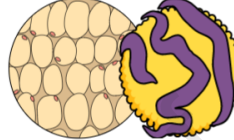
FINDINGS

ApoB is an important weight management target, especially with discordantly high ApoB.

Dietary fibre and saturated fat intake is associated with discordantly high ApoB.

ApoB has use as a lifestyle therapeutic and recommendation target, especially when discordant with LDL-C

The Role of Adiposity, Diet and Inflammation on the Discordance between LDL-C and Apolipoprotein B.

Richard J. Webb[a*], Mohsen Mazidi[b], Gregory Y. H. Lip[c,d], Andre P. Kengne[e], Maciej Banach[f,g], Ian G. Davies[h]

a. School of Health Sciences, Faculty of Science, Liverpool Hope University, Hope Park Campus, Taggart Avenue, Liverpool, L16 9JD, United Kingdom, Email: webbr1@hope.ac.uk

b. Department of Twin Research & Genetic Epidemiology, King's College London, 4th Floor, South Wing St Thomas', St Thomas', London, SE1 7EH, United Kingdom, Email: mohsen.mazidi@kcl.ac.uk

c. Liverpool Centre for Cardiovascular Science, University of Liverpool and Liverpool Heart & Chest Hospital, Liverpool L14 3PE, United Kingdom, Email: gregory.lip@liverpool.ac.uk

d. Aalborg Thrombosis Research Unit, Department of Clinical Medicine, Aalborg University, DK-9100 Aalborg, Denmark,

e. Non-Communicable Diseases Research Unit, South African Medical Research Council, Cape Town, South Africa, Email: Andre.Kengne@mrc.ac.za

f. Cardiovascular Research Centre, University of Zielona Gora, Zielona Gora, Poland,

g. Department of Hypertension, Medical University of Lodz (MUL), Lodz, Poland; E-mail: maciej.banach@umed.lodz.pl

h. School of Sports and Exercise Sciences, Faculty of Science, Liverpool John Moores University, IM Marsh Campus, Barkhill Road, Liverpool L17 6AF, United Kingdom, Email: I.G.Davies@ljmu.ac.uk

*To whom correspondence should be addressed and preprints be made: Email: webbr1@hope.ac.uk/mohsen.mazidi@kcl.ac.uk.

Word count: 4510 (Excluding abstract)

Number of figures and tables: Figures: 2, Tables: 2

Abbreviations: Structural equation models (SEMs); Mendelian randomisation (MR); saturated fatty acid (SFA); C-reactive protein (CRP); inverse variance weighted (IVW); atherosclerotic cardiovascular disease (ASCVD); National Health and Nutrition Examination Survey (NHANES); National Center for Health Statistics (NCHS); United States Department of Agriculture Automated Multiple-Pass Method (AMPM); waist circumference (WC); fasting blood glucose (FBG); homeostatic model assessment of insulin resistance (HOMA-IR); glycated haemoglobin (HbA1c); serum uric acid (SUA); random forest (RF); comparative fit index (CFI); Tucker–Lewis index (TLI); root mean square error of approximation (RMSEA); genome-wide association studies (GWAS); Global Lipid Genetics Consortium (GLGC); linkage disequilibrium (LD); weighted median (WM); instrument strength independent of direct effect (InSIDE); MR pleiotropy residual sum and outlier (MR-PRESSO); MR-Robust Adjusted Profile Score (RAPS); systolic blood pressure (SBP); diastolic blood pressure (DBP)

Abstract:

Background and Aims: While low-density lipoprotein cholesterol (LDL-C) is a good predictor of atherosclerotic cardiovascular disease, apolipoprotein B (ApoB) is superior when the two markers are discordant. We aimed to determine the impact of adiposity, diet and inflammation upon ApoB and LDL-C discordance.

Methods and Results: Machine learning (ML) and structural equation models (SEMs) were applied to the National Health and Nutrition Examination Survey to investigate cardiometabolic and dietary factors when LDL-C and ApoB are concordant/discordant. Mendelian randomisation (MR) determined whether adiposity and inflammation exposures were causal of elevated/decreased LDL-C and/or ApoB. ML showed body mass index (BMI), dietary saturated fatty acids (SFA), dietary fibre, serum C-reactive protein (CRP) and uric acid were the most strongly associated variables ($R^2 = 0.70$) in those with low LDL-C and high ApoB. SEMs revealed that fibre (b = -0.42, $p = 0.001$) and SFA (b = 0.28, $p = 0.014$) had a significant association with our outcome (joined effect of ApoB and LDL-C). BMI (b = 0.65, $p = 0.001$), fibre (b = -0.24, $p = 0.014$) and SFA (b = 0.26, $p = 0.032$) had significant associations with CRP. MR analysis showed genetically higher body fat percentage had a significant causal effect on ApoB (Inverse variance weighted (IVW) = Beta: 0.172, $p = 0.0001$) but not LDL-C (IVW = Beta: -0.006, $p = 0.845$).

Conclusion: Our data show increased discordance between ApoB and LDL-C is associated with cardiometabolic, clinical and dietary abnormalities and that body fat percentage is causal of elevated ApoB.

3

1. Introduction:

Low-density lipoprotein cholesterol (LDL-C) is a good predictor of atherosclerotic cardiovascular disease ASCVD, and the main target for pharmacological therapy [1]. However, there remains a residual cardiovascular risk, even after controlling for LDL-C in patients with metabolic syndrome and inflammation [1-5]. A common feature of all atherogenic lipoproteins is that they carry one molecule of apolipoprotein B (ApoB); an attribute which allows the molecule to be used as a measure of the amount of these particles [6]. Compared to LDL-C, elevated ApoB is a superior predictor of ASCVD risk when there is discordance between the two markers [7].

Recent studies have investigated ApoB with Mendelian randomisation (MR), which is a powerful method of inferring causality within an observational epidemiological context by using genetic variants as natural experiments [8, 9]. Moreover, MR benefits from being less susceptible to confounding and reverse causation [9]. When applying MR, ApoB containing particles are the main causal trait responsible for the aetiology of ASCVD [8, 10].

Discordantly high ApoB compared to LDL-C predominates in pro-inflammatory states, such as obesity and metabolic syndrome [11, 12]. Of note, pharmacotherapy for inflammation shows conflicting outcomes, due to different targeted pathways [13-15]. Combined with the degree of discordance between LDL-C and ApoB, this suggests that lowering LDL-C and inflammation are not always appropriate for addressing ASCVD, unless there is a concomitant reduction in ApoB. While weight loss studies show ApoB to be more closely related to improvements in adiposity [16], the causal links between body fat and ApoB have not been elucidated. Furthermore, there are no studies regarding the influence of lifestyle factors upon the degree of discordance of LDL-C with ApoB, despite nutritional factors strongly modulating lipoproteins and ASCVD risk [17-20].

Despite considerable evidence demonstrating the role of nutrition in ASCVD risk, contemporary analytical approaches can be applied to yield novel insights. For example, machine learning (ML) has recently gained attention due to its ability to elucidate unique relationships within large datasets [21]. In part, this is due to traditional regression techniques failing to coherently explain relationships between predictors and outcomes as these datasets often contain complex non-linear data with many predictors [22]. Methods to establish the magnitude of associations found within the data,

4

such as structural equation models (SEMs), which assess complex and multivariable relationships that benefit from multicollinearity, and the ability to elucidate complex networks, have recently been incorporated into cardiometabolic research [23, 24].

The current study employed two novel approaches to investigate the effects of obesity, inflammation, and dietary factors on ApoB and LDL-C and their discordance. First, MR was employed to investigate whether adiposity and sub-clinical inflammation exposures were causally linked to LDL-C and/or ApoB. Second, categories of LDL-C/ApoB discordance were created to determine the influence of adiposity, inflammation and diet data from the large-scale US National Health and Nutrition Examination Survey (NHANES). Specifically, ML and SEMs were combined to highlight unique predictors of interest and their magnitude of contribution.

2. Methods:

2.1 Study population:

This was a cross-sectional study (summarised in Figure 1) using data derived from the US National Health and Nutrition Examination Survey (NHANES). The National Center for Health Statistics (NCHS) Research Ethics Review Board approved the underlying protocol. Written informed consent was obtained from all participants and the study complied with the 1975 Declaration of Helsinki for medical research involving human subjects. The current study was based on the analysis of data for two 2-year NHANES survey cycles between 2005 and 2012, restricted to participants aged ≥ 18 years. Details on NHANES Laboratory/Medical Technologists Procedures and Anthropometry Procedures are described elsewhere [25]. A blood sample was drawn from the participant' antecubital vein. Details on laboratory-test details are available in the NHANES Laboratory/Medical Technologists Procedures Manual [25].

Details on recording dietary intake have been previously described [26]. Briefly, dietary intake was assessed via 24 h recall obtained by a trained interviewer, with the use of a computer-assisted dietary interview system with standardised probes using the United States Department of Agriculture Automated Multiple-Pass Method (AMPM) [26]. The AMPM is designed to enhance complete and accurate data collection while reducing respondent burden [26, 27]. The United States

Department of Agriculture (USDA) Food and Nutrient Database for Dietary Studies was used to

determine the nutrient content of food during the NHANES survey [28].

2.2 Statistical analysis:

High/low levels of LDL-C and ApoB were defined by cut-off values of 160 mg/dL and 130

mg/dL respectively [2], which resulted in four concordant/discordant categories. Further analyses

were conducted according to the guidelines of the Centers for Disease Control and Prevention for

analysis of the NHANES dataset, accounting for masked variance and using their suggested weighting

methodology [29]. Continuous and categorical demographic variables were compared across the four

groups using analysis of variance (ANOVA) and Chi-square tests respectively.

2.3 Machine Learning:

We used ML to assess which features [(body mass index (BMI), waist circumference (WC),

fasting blood glucose (FBG), plasma insulin, homeostatic model assessment of insulin resistance

(HOMA-IR), glycated haemoglobin (HbA_{1c}), alanine aminotransferase, aspartate aminotransferase,

gamma glutamyl transferase, total fat, monounsaturated fatty acids (MUFA), polyunsaturated fatty

acids (PUFA), saturated fatty acids (SFA), protein, carbohydrate, fibre, total sugar, serum uric acid

(SUA), serum C-reactive protein (CRP) and total bilirubin)] influence LDL-C and ApoB discordance.

We hypothesised that each independent factor may have a variable effect on the level of LDL-C and

ApoB. Therefore, we implemented our model for each of the four groups separately to reveal

predictors of our outcome (i.e. a joint effect of both ApoB and LDL-C was produced by dimension

reduction method, principal component analysis, will be referred to herein as their 'joined effect'). A

random forest (RF) model was applied with cross validation. This method fits many classification

trees to a data set, then combines the predictions from all trees to present a final predictive model that

ranks variables by their predictive power. However, this model does not provide mechanistic insight

and may mask variable interaction and nonlinearity. For the evaluation of our models we have used R^2

and Q^2 (an estimate of the predictive ability of the model calculated by cross-validation). A negative

Q^2 means the model is not at all predictive.

6

2.4 Structural equation modeling (SEM):

We used structural equation modeling (SEM) to test the overall model fit and relationships between sets of variables which were selected from machine learning to understand the underlying relationship of the combined LDL-C/ApoB joined effect category (for each group separately). SEMs are able to test the fit of the defined model based on the observed covariance between the variables. We fitted our model under a maximum likelihood framework using covariance matrices [30]. All continuous variables were standardised by rank-normal transformed (mean 0, SD 1) by age and sex (and by medication history). Relative model fit was assessed using the comparative fit index (CFI) and the Tucker–Lewis index (TLI), with values ranging from 0 (no fit) to 1 (perfect fit); a model with a 'good' fit typically requires both indices to exceed 0.95. Absolute fit was assessed using the root mean square error of approximation (RMSEA). This ranges from 0 to 1, with 0 indicating a perfect fit [30]. A poorly fitting model is typically defined by RMSEA > 0.06 [31]. CFI, TLI and RMSEA were not used to formally determine adequacy of fit, as their use in this context is controversial and there is limited consensus on appropriate cut-off values because each index is affected differently by degrees of freedom, model complexity and sample size; however, it is standard practice to report these along with the χ^2. To overcome this, we formally tested the model fit by comparing the χ^2 of the tested model with χ^2 values obtained from variable-randomised null models with identical structures (in other words, the variables were randomly assigned to other nodes in the same structural equation model definition) and applied to the respective covariance matrix used for the tested model. This process was iterated 10,000 times and we reported the mean of χ^2 values for the real model and null model. A two-sided $p < 0.05$ was used to characterise significant results. Statistical analysis for the SEMS was performed in the 'Lavaan' package for the R environment for statistical computing v 3.5.1 (R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/).

2.5 Mendelian randomisation:

We employed MR to determine whether the relationships in our SEM are causal. We have chosen body fat percentage and CRP as they were affecting the outcome in most of the models.

7

Furthermore, we had data to run the MR for the body fat percentage and CRP. Both are modifiable

risk factors which can be used for clinical advice and to inform future randomised control trials

targeting the modification of both LDL-C and ApoB. We had single nucleotide polymorphism (SNP)

instruments for both body fat percentage and CRP.

2.5.1 Genetic predictors:

Genetic associations for body fat percentage and CRP were obtained from UK Biobank data and

other genome-wide association studies (GWAS). More details can be found elsewhere [32, 33].

Genetic associations with fasting ApoB (quantified by nuclear magnetic resonance) were obtained

from the largest available extensively genotyped study (among 24,925 adults). Again, more details

can be found elsewhere [34]. We retrieved summary data for the association between SNPs and

circulating fasting LDL-C from the Global Lipid Genetics Consortium (GLGC) (188,577 adult

samples of European ancestry). They included rigorous quality control, imputation to the 1000

Genomes Project panel and adjustments for age and population structure. Persons of European

ancestry from 47 studies genotyped with different genome-wide association study arrays (n = 94,595)

or on the Metabochip array (n = 93,982) with imputation to the 1000 Genomes Project reference were

studied. In most included studies, blood lipid concentrations had been measured after > 8 hours of

fasting. Participants on lipid lowering medications were excluded. Traits were adjusted for age, age-

squared, sex and principal components, as well as quantile-normalized within each cohort. For genetic

association analysis by linear regression, lipid levels were inverse normal-transformed and cohort-

wise results combined in fixed effect meta-analysis.

If a SNP was unavailable for the outcome GWAS summary statistics, we identified proxy SNPs

with a minimum linkage disequilibrium (LD) $R^2 = 0.8$. To minimize bias in effect estimates induced

by correlation between SNPs, we restricted our genetic instrument to independent SNPs not in linkage

disequilibrium ($p = 0.0001$). We refer to a set of SNPs that proxy serum Lead as "genetic

instruments."

2.5.2 Mendelian Randomisation Statistics:

We combined the effect of instruments using the inverse variance weighted (IVW) method. Heterogeneity was assessed using Q value for IVW. To address the potential effect of pleiotropic variants on the final effect estimate, we performed sensitivity analysis including weighted median (WM) and MR-Egger. Sensitivity analysis was conducted using the leave-one-out method to identify instruments that might drive the MR results. The WM estimate provides correct estimates if SNPs accounting for $\geq 50\%$ of the weight are valid instruments. Inverse variance is used to weight the variants and bootstrapping is applied to estimate the CIs [35]. MR-Egger can make estimates even under the assumption that all SNPs are invalid instruments, as long as the assumption of instrument strength independent of direct effect (InSIDE) is satisfied [35]; however, the InSIDE assumption cannot be easily verified. Average directional pleiotropy across genetic variants was assessed from the *p* value of the intercept term from MR-Egger [35]. Causal estimates in MR-Egger are less precise than those obtained by using IVW MR [36]. Analysis using MR-Egger has a lower false-positive rate, but a higher false-negative rate, than IVW i.e., it has a lower statistical power [36].

Heterogeneity between individual genetic variant estimates was assessed using the Q′ heterogeneity statistic [37]. The Q′ statistic uses modified 2nd order weights that are a derivation of a Taylor series expansion, considering the uncertainty in both numerator and denominator of the instrumental variable ratio [37].

2.5.3 Sensitivity analysis:

As sensitivity analysis, we used MR-Egger and MR pleiotropy residual sum and outlier (MR-PRESSO) test [38]. MR-Egger and MR-PRESSO may provide correct estimates as long as the instrument strength independent of direct effect assumption is satisfied [38]. MR-Egger can be imprecise, particularly if the associations for SNPs on exposure are similar, or the number of genetic instruments is low [38]. A non-null MR-Egger intercept suggests that the IVW estimate is invalid. MR-Egger does not explicitly identify outliers. MR-PRESSO detects, and if necessary, corrects for potentially pleiotropic outliers [38]. The MR-PRESSO framework detects effect estimates that are outliers and removes them from the analysis by regressing the variant-outcome associations on variant-exposure associations. A global heterogeneity test is then implemented to compare the

9

observed distance between residual sums of squares of all variants to the regression line with the distance expected under the null hypothesis of no pleiotropy [39]. Furthermore, MR-Robust Adjusted Profile Score (RAPS) was applied, which can correct for pleiotropy using robust adjusted profile scores. We considered causal estimates that agreed in direction and magnitude across MR methods, passed nominal significance in IVW MR, and did not show evidence of bias from horizontal pleiotropy using heterogeneity tests. Statistical analysis was for MR performed in the 'TwoSampleMR' package for the R environment for statistical computing v 3.5.1 (R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/).

3. Results:

3.1 Demographics and Clinical Characteristics:

Table 1 shows the demographic and clinical characteristics of the study population based on LDL-C and ApoB concordant/discordant categories. Group 1 (Low LDL-C, Low ApoB) had the highest population (n=12,384), followed by group 4 (n=885), group 2 (n=711) and group 3 (n=285). Of the participants 48.7% were male. The mean age was 47.6 years overall, 47.8 years in men and 47.3 years in women ($p = 0.098$). Overall mean body mass index (BMI) and waist circumference (WC) were 28.7 kg/m^2 and 99 cm respectively. Descriptive variables categorised by high/low LDL-C and ApoB are shown in **Table 1**. Significant differences were apparent for all continuous and categorical demographic variables across the four groups (all $p < 0.001$). For instance, subjects in the third group (Low LDL-C, High ApoB) had significantly higher levels of adiposity, BMI (30.9 ± 0.9 kg/m^2) and WC (106 ± 2 cm) compared to the other groups (both $p < 0.001$). With regard to insulin and glucose parameters, fasting blood glucose (FBG) (118 ± 10 mg/dL), insulin (16.77 ± 1.68 µU/mL) and HOMA-IR (5.21 ± 0.86), the third group (Low LDL-C, High ApoB) had the highest cardio-metabolic risk profile when compared to the other groups (all $p < 0.001$, **Table 1**).

3.2 Machine Learning:

Our analysis revealed that BMI, FBG, dietary fat and serum uric acid were the most strongly associated variables ($R^2 = 0.64$) in the first group (Low LDL-C, Low ApoB). For the second group

(High LDL-C, Low ApoB) WC, FBG, gamma glutamyl transferase, dietary SFA, dietary PUFA, dietary fibre, total serum bilirubin and serum CRP were the most strongly associated predictors ($R^2 = 0.72$) for our outcome (joined effect of ApoB and LDL-C). With respect to the third group (Low LDL-C, High ApoB), BMI, dietary SFA, dietary fibre, serum CRP and serum uric acid revealed the strongest relationships ($R^2 = 0.70$). Finally, within the last group (High LDL-C, High ApoB) BMI, dietary fat, dietary carbohydrate, total serum bilirubin and serum uric acid were the most important associated variables ($R^2 = 0.61$).

3.3 Structural Equation Models:

To determine the effect magnitude for each of the predictors resulting from our ML analysis of the four LDL-C and ApoB concordant/discordant categories we implemented a structural equation model (SEM) for each group (**Supplemental Tables S1 to S4**).

The SEM applied to the first group (Low LDL-C, Low ApoB) showed that SUA was the only variable which had an interdependent significant association (b = -0.15, $p = 0.043$) with our outcome (joined effect of ApoB and LDL-C). Additionally, both SUA (b = -0.28, $p = 0.001$) and BMI (b = 0.18, $p = 0.012$) had a significant association with FBG (**Supplemental Table S1**). Regarding the second group (High LDL-C, Low ApoB), our SEM showed that CRP (b = 0.96, $p = 0.001$) and fibre intake (b = -0.21, $p = 0.042$) both had a significant independent association with our outcome. Furthermore, fibre intake (b = -0.31, $p = 0.025$), SFA intake (b = 0.56, $p = 0.001$) and WC (b = 0.96, $p = 0.001$) had a significant relationship with CRP (**Supplemental Table S2**). The SEM revealed that fibre intake (b = -0.42, $p = 0.001$) and SFA intake (b = 0.28, $p = 0.014$) had a significant association with our outcome in the third group (Low LDL-C, High ApoB). The effect estimates are presented in **Figure 2**. Furthermore, BMI (b = 0.65, $p = 0.001$), fibre intake (b = -0.24, $p = 0.014$) and SFA intake (b = 0.26, $p = 0.032$) also had a significant association with CRP (**Supplemental Table S3**). With respect to the fourth group (High LDL-C, High ApoB), our SEM revealed that SUA (b = 0.23, $p = 0.001$) and BMI (b = 0.26, $p = 0.001$) both had a significant relationship with our outcome. Furthermore, BMI was also significantly associated with bilirubin (b = -0.19, $p = 0.001$) and SUA (b

11

= 0.55, *p* = 0.001) **(Supplemental Table S4)**. All the models demonstrated a good fit (*Chi-square*: 14.1, CFI: 0.995, RMSEA: 0.050, TLI: 0.958).

3.4 Mendelian Randomisation:

The instruments have F-statistics ranging from 326 to 425, making significant bias from the use of weak instruments unlikely. The results, expressed as beta-coefficients per 1 standard deviation (SD) increase in outcomes, are presented in **Table 2**.

Genetically higher body fat percentage had a significant effect on ApoB (IVW = Beta: 0.172, *p* = 0.0001, **Table 2, Supplemental Figure S1**) but not LDL-C levels (IVW = Beta: -0.006, *p* = 0.845, **Table 2, Supplemental Figure S2**). Higher CRP levels had no significant effect on ApoB (IVW = Beta: 0.032, *p* = 0.502, **Table 2, Supplemental Figure S3**) or LDL-C levels (IVW = Beta: -0.046, *p* = 0.247, **Table 2, Supplemental Figure S4**).

Heterogeneity results and pleiotropy bias are also shown in **Table 2**. Estimation is based on both MR Egger and IVW indicted chance of heterogeneity for all of our predictions (all IVW *p* < 0.0023, all MR Egger *p* < 0.0020). We performed MR-PRESSO (to detect outlier SNPs and estimate corrected effects) which revealed the effect of body fat percentage on ApoB (Beta: 0.190, *p* = 0.0003, Table 2) and LDL-C (Beta: -0.001, *p* = 0.971, **Table 2**) and the further impact of CRP on ApoB (Beta: -0.023, *p* = 0.392, **Table 2**) and LDL-C (Beta: -0.019, *p*= 0.347, **Table 2**). The horizontal pleiotropy test, with very negligible Egger regression intercept, also indicated a low likelihood of pleiotropy for all our estimations (all *p* > 0.111). The results of the MR-RAPS were identical with the IVW estimates, highlighting again a low likelihood of pleiotropy. The results of the leave-one-out method demonstrated that the links were not driven by single SNPs.

4. Discussion:

In this study, we aimed to investigate the role of cardiometabolic and dietary factors in relation to LDL-C/ApoB discordance. We found that those belonging to the third group (Low LDL-C, High ApoB) had the worst profile of cardiometabolic risk markers when compared to the other groups. Our novel approach with ML and SEM revealed for the first time that predictors of the

combined measure of LDL-C and ApoB are different for each group. These differences were implicated by markers of inflammation, adiposity, and dietary intake of fibre and SFA of varying magnitudes and significance within each group. Indeed, the use of ML followed by SEM revealed a significant relationship of fibre and SFA intake within the Low LDL-C/High ApoB group, whereas the High LDL-C/Low ApoB group showed both CRP and fibre intake had a significant independent association with joined effect of ApoB and LDL-C.

Analysis using MR revealed that body fat percentage was causal of ApoB but not LDL-C and the inflammatory marker, CRP, had no causal relationship with neither ApoB nor LDL-C. This is the first study where a causal role between adiposity and ApoB has been demonstrated. Relationships found by others regarding body fat, BMI, WC and other measures of adiposity with ApoB, together with convincing mechanistic evidence, further support our findings [16, 40, 41]. It has long been established that adiposity, especially visceral adipose tissue and insulin resistance, lead to excess secretion of fatty acids (FA) via the suppression of hormone sensitive lipase [42]. This causes FA to accumulate in the liver, resulting in the synthesis of ApoB100 and large VLDL [42]. Furthermore, the clearance of ApoB in obesity is compromised due to the underproduction of lipoprotein lipase, leading to impaired hydrolysis of VLDL and increased plasma residence time [43].

With regards to LDL-C and ApoB discordance, the present study revealed distinct categories based upon the levels of these markers which reflect the heterogeneous variance found in the population. These discordant categories represented substantially less than the ~20% prevalence cited by other studies; however, previous research has employed different ApoB and LDL-C cut-offs, or analysed at risk populations (e.g. T2D, metabolic syndrome) [44-46]. We applied cut offs of 130 mg/dL and 160 mg/dL for ApoB and LDL-C respectively, levels agreed upon by expert consensus [2], in an otherwise healthy population from the NHANES database. Despite these more stringent cut offs, discordance between LDL-C and ApoB remained and differences between the four groups were revealed which aligned with those previously reported [47]. Supporting our finding of the causal role of body fat percentage on ApoB, we demonstrated significant differences with BMI and WC between groups (Table 1). Indeed, BMI reaching obesity was observed in the high ApoB groups, and levels were practically identical between these two groups. However, WC was highest in Group 3 (Low

LDL-C, High ApoB), suggesting higher visceral obesity and supporting previous literature as the group with the most disrupted metabolism, and therefore highest cardiometabolic risk [48, 49]. Other markers in Group 3 which corroborate this include higher insulin resistance, systolic blood pressure, lower HDL-C, and higher triglycerides.

It is known that CRP and SUA are intimately associated with inflammatory processes, which are predictive of ASCVD risk [50, 51]. This inflammation is thought to partly result from LDL particles stimulating endothelial cells which increase the production of CRP, which in turn stimulates the release of lectin-like oxidized LDL receptor 1 from macrophages [52]. This further increases the uptake of LDL, creating a vicious cycle [52]. Furthermore, SUA has also been shown to directly regulate proinflammatory pathways in vascular smooth muscle cells, further contributing towards the nefarious progression of ASCVD [53]. Our results are in agreement with significant relationships revealed between both CRP and SUA and ApoB/LDL-C within Groups 2 and 4 (High LDL-C, Low ApoB and High LDL-C, High ApoB respectively). Moreover, a negative association between SUA in Group 1 (Low LDL-C, Low ApoB) was also shown, further emphasising the potential prognostic value of these markers in ASCVD.

The influence of dietary factors showed small but significant differences for total fat, SFA, PUFA, and MUFA, but these were not clinically significant and were largely within recommended guidelines (Table 1) [54]. Only total carbohydrate and total sugar intake were significantly higher in Group 3 (Low LDL-C, High ApoB). This is supported by literature showing carbohydrate, especially refined carbohydrate, increases plasma TG and small dense LDL, lowers HDL-C, and negatively impacts markers of glucose metabolism and inflammation [55]. The significantly lower LDL-C/ApoB and higher TG/HDL-C ratios in this group is suggestive of a 'Pattern B profile', which is predominant in insulin resistance and low-grade inflammatory states and correlated with SUA [56]. Krauss et al. [55] and others have shown a consistent improvement of this pattern with a lower carbohydrate approach [57, 58]. Larger scale studies, such as the PURE study, have also demonstrated that a lower carbohydrate diet results in a more favourable overall blood lipid profile compared to that of a high carbohydrate and low-fat diet [59]. However, one of the main criticisms of this strategy is an overall increase in LDL-C and debate continues regarding the promotion of such diets [60, 61]. The few

14

lower carbohydrate studies that investigated atherogenic lipoproteins (including ApoB) show a high

degree of variance, with or without weight loss, suggesting dietary factors such as saturated fat might

influence ApoB [57, 62]. Findings from Furtado et al. demonstrate this by showing that dietary

patterns which are high in saturated fat are associated with higher levels of ApoB than those which

emphasise carbohydrates, unsaturated fat or protein [63]. This was partially supported by our data,

showing a weak but significant correlation with SFA and ApoB/LDL-C in Group 3 (Low LDL-C,

High ApoB) but not in Group 4 (High LDL-C, High ApoB). Indeed, Group 4 had significantly lower

total and SFA intake, incongruent with the diet-heart hypothesis [64], suggesting other factors such as

various SNPs may contribute [65].

Despite the differences between groups being small and all groups consuming less than

recommendations, there was a highly significant and moderate negative correlation with dietary fibre

and ApoB and LDL-C in Group 3 (Low LDL-C, High ApoB). Furthermore, the high intake of overall

carbohydrate and low intake of fibre in this group may suggest that intake may be comprised

predominantly of refined carbohydrate. These findings are in alignment with limited human

randomised controlled trials, which have shown that in individuals with dyslipidaemia the

consumption of soluble fibre results in decreased levels of ApoB via the reduced reabsorption of bile

acids and increased excretion of cholesterol [16].

4.1 Limitations:

Our study has some limitations. Firstly, while we consider using consensus cut-off points for

LDL-C and ApoB, other studies use median values which may reveal further differences between

groups. Second, MR should ideally be performed in different ethnicities to ensure validity of the

findings. Third, it would have been preferential to have GWAS data for the discordant groups, but this

was not possible as no database exists which contains this data. Fourth, the authors did not have

clinical endpoint data available for the participants, such as CVD event or mortality, as the NHANES

database does not contain a large enough sample. Fifth, although the predictive value of the ratio of

ApoB to anti-atherogenic apolipoprotein A1 (ApoA1) and its strong relationship with body fat

distribution is well-documented, NHANES did not measure ApoA1 making its ratio with ApoB

impossible to calculate and utilise in our study [66]. Finally, there are inherent and well-documented limitations with 24-hour recall dietary assessment data, including recall bias which may lead to under and/or over reporting [67].

4.2 Conclusions:

In conclusion, our findings reveal several novel findings of significant importance for ASCVD risk that should guide future recommendations. First, we show a causal relationship between body fat percentage and ApoB suggesting weight management as a powerful strategy to reduce ASCVD risk. Second, we reveal that the subgroup pertaining to discordantly high ApoB in relation to LDL-C is associated with several cardiometabolic, clinical, and anthropometric abnormalities and poor dietary intake. Finally, our data supports the use of ApoB as a lifestyle therapeutic and target for recommendations rather than LDL-C *per se*, especially when the two measures are discordant.

References:

[1] Boren J, Chapman MJ, Krauss RM, Packard CJ, Bentzon JF, Binder CJ, et al. Low-density lipoproteins cause atherosclerotic cardiovascular disease: pathophysiological, genetic, and therapeutic insights: a consensus statement from the European Atherosclerosis Society Consensus Panel. Eur Heart J. 2020.

[2] Grundy SM, Stone NJ, Bailey AL, Beam C, Birtcher KK, Blumenthal RS, et al. 2018 AHA/ACC/AACVPR/AAPA/ABC/ACPM/ADA/AGS/APhA/ASPC/NLA/PCNA Guideline on the Management of Blood Cholesterol: Executive Summary: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. J Am Coll Cardiol. 2019;73:3168-209.

[3] Langlois MR, Nordestgaard BG, Langsted A, Chapman MJ, Aakre KM, Baum H, et al. Quantifying atherogenic lipoproteins for lipid-lowering strategies: consensus-based recommendations from EAS and EFLM. Clin Chem Lab Med. 2020;58:496-517.

[4] Patel KV, Pandey A, de Lemos JA. Conceptual Framework for Addressing Residual Atherosclerotic Cardiovascular Disease Risk in the Era of Precision Medicine. Circulation. 2018;137:2551-3.

[5] Vallejo-Vaz AJ, Corral P, Schreier L, Ray KK. Triglycerides and residual risk. Current Opinion in Endocrinology, Diabetes and Obesity. 2020;27:95–103.

[6] Catapano AL. Atherogenic lipoproteins as treatment targets. Nature Reviews Cardiology. 2018;15:75-6.

[7] Sniderman AD, Thanassoulis G, Glavinovic T, Navar AM, Pencina M, Catapano A, et al. Apolipoprotein B Particles and Cardiovascular Disease: A Narrative Review. JAMA Cardiol. 2019;4:1287-95.

[8] Ference BA, Kastelein JJP, Ray KK, Ginsberg HN, Chapman MJ, Packard CJ, et al. Association of Triglyceride-Lowering LPL Variants and LDL-C–Lowering LDLR Variants With Risk of Coronary Heart Disease. JAMA. 2019;321:364-73.

[9] Davies NM, Holmes MV, Smith GD. Reading Mendelian randomisation studies: a guide, glossary, and checklist for clinicians. BMJ. 2018;362.

[10] Richardson TG, Sanderson E, Palmer TM, Ala-Korpela M, Ference BA, Davey Smith G, et al. Evaluating the relationship between circulating lipoprotein lipids and apolipoproteins with risk of

coronary heart disease: A multivariable Mendelian randomisation analysis. PLoS Med. 2020;17:e1003062.

[11] Cao J, Nomura SO, Steffen BT, Guan W, Remaley AT, Karger AB, et al. Apolipoprotein B discordance with low-density lipoprotein cholesterol and non-high-density lipoprotein cholesterol in relation to coronary artery calcification in the Multi-Ethnic Study of Atherosclerosis (MESA). J Clin Lipidol. 2020;14:109-21.e5.

[12] Laclaustra M, Lopez-Garcia E, Civeira F, Garcia-Esquinas E, Graciani A, Guallar-Castillon P, et al. LDL Cholesterol Rises With BMI Only in Lean Individuals: Cross-sectional U.S. and Spanish Representative Data. Diabetes Care. 2018;41:2195-201.

[13] Ridker PM, Everett BM, Thuren T, MacFadyen JG, Chang WH, Ballantyne C, et al. Antiinflammatory Therapy with Canakinumab for Atherosclerotic Disease. N Engl J Med. 2017;377:1119-31.

[14] Ridker PM, Everett BM, Pradhan A, MacFadyen JG, Solomon DH, Zaharris E, et al. Low-Dose Methotrexate for the Prevention of Atherosclerotic Events. N Engl J Med. 2018;380:752-62.

[15] Ridker PM, Libby P, MacFadyen JG, Thuren T, Ballantyne C, Fonseca F, et al. Modulation of the interleukin-6 signalling pathway and incidence rates of atherosclerotic events and all-cause mortality: analyses from the Canakinumab Anti-Inflammatory Thrombosis Outcomes Study (CANTOS). Eur Heart J. 2018;39:3499-507.

[16] Lamantia V, Sniderman A, Faraj M. Nutritional management of hyperapoB. Nutr Res Rev. 2016;29:202–33.

[17] Clifton PM, Keogh JB. A systematic review of the effect of dietary saturated and polyunsaturated fat on heart disease. Nutr Metab Cardiovasc Dis. 2017;27:1060-80.

[18] Livingstone KM, Lovegrove JA, Givens DI. The impact of substituting SFA in dairy products with MUFA or PUFA on CVD risk: evidence from human intervention studies. Nutr Res Rev. 2012;25:193-206.

[19] Position of the American Dietetic Association and Dietitians of Canada: Dietary Fatty Acids. J Am Diet Assoc. 2007;107:1599.e1-.e15.

[20] Perk J, De Backer G, Gohlke H, Graham I, Reiner Ž, Verschuren M, et al. European Guidelines on cardiovascular disease prevention in clinical practice (version 2012): The Fifth Joint Task Force of the European Society of Cardiology and Other Societies on Cardiovascular Disease Prevention in Clinical Practice (constituted by representatives of nine societies and by invited experts)Developed with the special contribution of the European Association for Cardiovascular Prevention & Rehabilitation (EACPR)†. Eur Heart J. 2012;33:1635-701.

[21] Rigdon J, Basu S. Machine learning with sparse nutrition data to improve cardiovascular mortality risk prediction in the USA using nationally randomly sampled data. BMJ Open. 2019;9:e032703.

[22] Sidey-Gibbons JAM, Sidey-Gibbons CJ. Machine learning in medicine: a practical introduction. BMC Med Res Methodol. 2019;19:64-.

[23] Khodarahmi M, Asghari-Jafarabadi M, Abbasalizad Farhangi M. A structural equation modeling approach for the association of a healthy eating index with metabolic syndrome and cardio-metabolic risk factors among obese individuals. PLoS One. 2019;14:e0219193.

[24] Yamada T, Ogawa K, Tanaka TD, Nagoshi T, Minai K, Ogawa T, et al. Increase in oxidized low-density lipoprotein level according to hyperglycemia in patients with cardiovascular disease: A study by structure equation modeling. Diabetes Res Clin Pract. 2020;161.

[25] CDC. National Health and Nutrition Survey - Laboratory Procedures Manual. In: CDC, editor.2011.

[26] Ahluwalia N, Dwyer J, Terry A, Moshfegh A, Johnson C. Update on NHANES Dietary Data: Focus on Collection, Release, Analytical Considerations, and Uses to Inform Public Policy. Adv Nutr. 2016;7:121-34.

[27] Moshfegh AJ, Rhodes DG, Baer DJ, Murayi T, Clemens JC, Rumpler WV, et al. The US Department of Agriculture Automated Multiple-Pass Method reduces bias in the collection of energy intakes. Am J Clin Nutr. 2008;88:324-32.

[28] U.S. Department of Agriculture ARS. USDA Food and Nutrient Database for Dietary Studies. 2019.

[29] Statistics. NCfH. Survey Methods and Analytic Guidelines. 2020.

[30] Bentler PM. Comparative fit indexes in structural models. Psychol Bull. 1990;107:238-46.

[31] Tucker LR, Lewis C. A reliability coefficient for maximum likelihood factor analysis. Psychometrika. 1973;38:1-10.

[32] Dehghan A, Dupuis J, Barbalic M, Bis JC, Eiriksdottir G, Lu C, et al. Meta-analysis of genome-wide association studies in >80 000 subjects identifies multiple loci for C-reactive protein levels. Circulation. 2011;123:731-8.

[33] Kocarnik JM, Pendergrass SA, Carty CL, Pankow JS, Schumacher FR, Cheng I, et al. Multiancestral analysis of inflammation-related genetic variants and C-reactive protein in the population architecture using genomics and epidemiology study. Circ Cardiovasc Genet. 2014;7:178-88.

[34] Kettunen J, Demirkan A, Würtz P, Draisma HHM, Haller T, Rawal R, et al. Genome-wide study for circulating metabolites identifies 62 loci and reveals novel systemic effects of LPA. Nat Commun. 2016;7:11122.

[35] Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator. Genet Epidemiol. 2016;40:304-14.

[36] Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. Int J Epidemiol. 2015;44:512-25.

[37] Burgess S, Bowden J, Fall T, Ingelsson E, Thompson SG. Sensitivity Analyses for Robust Causal Inference from Mendelian Randomization Analyses with Multiple Genetic Variants. Epidemiology. 2017;28:30-42.

[38] Verbanck M, Chen CY, Neale B, Do R. Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. Nat Genet. 2018;50:693-8.

[39] Palmer TM, Lawlor DA, Harbord RM, Sheehan NA, Tobias JH, Timpson NJ, et al. Using multiple genetic variants as instrumental variables for modifiable risk factors. Stat Methods Med Res. 2012;21:223-42.

[40] Dansinger ML, Williams PT, Superko HR, Schaefer EJ. Effects of weight change on apolipoprotein B-containing emerging atherosclerotic cardiovascular disease (ASCVD) risk factors. Lipids Health Dis. 2019;18:154.

[41] Mittendorfer B, Yoshino M, Patterson BW, Klein S. VLDL Triglyceride Kinetics in Lean, Overweight, and Obese Men and Women. The Journal of Clinical Endocrinology and Metabolism. 2016;101:4151-60.

[42] Vergès B. Pathophysiology of diabetic dyslipidaemia: where are we? Diabetologia. 2015;58:886–99.

[43] Björnson E, Adiels M, Taskinen M-R, Borén J. Kinetics of plasma triglycerides in abdominal obesity. Curr Opin Lipidol. 2017;28:11-8.

[44] Otvos JD, Mora S, Shalaurova I, Greenland P, Mackey RH, Goff DC. Clinical implications of discordance between low-density lipoprotein cholesterol and particle number. J Clin Lipidol. 2011;5:105–13.

[45] Tehrani DM, Zhao Y, Blaha MJ, Mora S, Mackey RH, Michos ED, et al. Discordance of Low-Density Lipoprotein and High-Density Lipoprotein Cholesterol Particle Versus Cholesterol Concentration for the Prediction of Cardiovascular Disease in Patients With Metabolic Syndrome and Diabetes Mellitus (from the Multi-Ethnic Study of Atherosclerosis [MESA]). The American journal of cardiology. 2016;117:1921–7.

[46] Cromwell WC, Otvos JD. Heterogeneity of low-density lipoprotein particle number in patients with type 2 diabetes mellitus and low-density lipoprotein cholesterol <100 mg/dl. The American Journal of Cardiology. 2006;98:1599-602.

[47] Otvos JD, Mora S, Shalaurova I, Greenland P, Mackey RH, Goff DC. Clinical Implications of Discordance Between LDL Cholesterol and LDL Particle Number. J Clin Lipidol. 2011;5:105-13.

[48] Després JP, Lemieux I, Bergeron J, Pibarot P, Mathieu P, Larose E, et al. Abdominal obesity and the metabolic syndrome: contribution to global cardiometabolic risk. Arterioscler Thromb Vasc Biol. 2008;28:1039-49.

[49] Neeland IJ, Ross R, Després J-P, Matsuzawa Y, Yamashita S, Shai I, et al. Visceral and ectopic fat, atherosclerosis, and cardiometabolic disease: a position statement. The Lancet Diabetes & Endocrinology. 2019;7:715-25.

[50] Ridker PM, Stampfer MJ, Rifai N. Novel Risk Factors for Systemic Atherosclerosis: A Comparison of C-Reactive Protein, Fibrinogen, Homocysteine, Lipoprotein(a), and Standard Cholesterol Screening as Predictors of Peripheral Arterial Disease. JAMA. 2001;285:2481-5.

[51] Kushiyama A, Nakatsu Y, Matsunaga Y, Yamamotoya T, Mori K, Ueda K, et al. Role of Uric Acid Metabolism-Related Inflammation in the Pathogenesis of Metabolic Syndrome Components Such as Atherosclerosis and Nonalcoholic Steatohepatitis. Mediators Inflamm2016.

[52] Lin G-M, Liu K, Colangelo LA, Lakoski SG, Tracy RP, Greenland P. Low-Density Lipoprotein Cholesterol Concentrations and Association of High-Sensitivity C-Reactive Protein Concentrations With Incident Coronary Heart Disease in the Multi-Ethnic Study of Atherosclerosis. Am J Epidemiol. 2016;183:46-52.

[53] Kanellis J, Watanabe S, Li JH, Kang DH, Li P, Nakagawa T, et al. Uric acid stimulates monocyte chemoattractant protein-1 production in vascular smooth muscle cells via mitogen-activated protein kinase and cyclooxygenase-2. Hypertension (Dallas, Tex: 1979). 2003;41:1287-93.

[54] USDA. Dietary Guidelines for Americans: Eighth Edition. 8th edition ed. https://health.gov/our-work/food-nutrition/2015-2020-dietary-guidelines/guidelines/2015.

[55] Krauss RM, Blanche PJ, Rawlings RS, Fernstrom HS, Williams PT. Separate effects of reduced carbohydrate intake and weight loss on atherogenic dyslipidemia. Am J Clin Nutr. 2006;83:1025-31; quiz 205.

[56] Peng T-C, Wang C-C, Kao T-W, Chan JY-H, Yang Y-H, Chang Y-W, et al. Relationship between Hyperuricemia and Lipid Profiles in US Adults. BioMed Research International2015.

[57] Wood RJ, Volek JS, Liu Y, Shachter NS, Contois JH, Fernandez ML. Carbohydrate Restriction Alters Lipoprotein Metabolism by Modifying VLDL, LDL, and HDL Subfraction Distribution and Size in Overweight Men. The Journal of Nutrition. 2006;136:384-9.

[58] Forouhi NG, Krauss RM, Taubes G, Willett W. Dietary fat and cardiometabolic health: evidence, controversies, and consensus for guidance. BMJ. 2018;361.

[59] Mente A, Dehghan M, Rangarajan S, McQueen M, Dagenais G, Wielgosz A, et al. Association of dietary nutrients with blood lipids and blood pressure in 18 countries: a cross-sectional analysis from the PURE study. The Lancet Diabetes & Endocrinology. 2017;5:774-87.

[60] Gjuladin-Hellon T, Davies IG, Penson P, Amiri Baghbadorani R. Effects of carbohydrate-restricted diets on low-density lipoprotein cholesterol levels in overweight and obese adults: a systematic review and meta-analysis. Nutr Rev. 2019;77:161-80.

[61] Mansoor N, Vinknes KJ, Veierød MB, Retterstøl K. Effects of low-carbohydrate diets v. low-fat diets on body weight and cardiovascular risk factors: a meta-analysis of randomised controlled trials. The British journal of nutrition. 2016;115:466–79.

[62] Chiu S, Williams PT, Krauss RM. Effects of a very high saturated fat diet on LDL particles in adults with atherogenic dyslipidemia: A randomized controlled trial. PLoS One. 2017;12:e0170664+.

[63] Furtado JD, Campos H, Appel LJ, Miller ER, Laranjo N, Carey VJ, et al. Effect of protein, unsaturated fat, and carbohydrate intakes on plasma apolipoprotein B and VLDL and LDL containing apolipoprotein C-III: results from the OmniHeart Trial. The American journal of clinical nutrition. 2008;87:1623-30.

[64] KEYS A. Human Atherosclerosis and the Diet. Circulation. 1952;5:115-8.

[65] Walker CG, Loos RJF, Olson AD, Frost GS, Griffin BA, Lovegrove JA, et al. Genetic predisposition influences plasma lipids of participants on habitual diet, but not the response to reductions in dietary intake of saturated fatty acids. Atherosclerosis. 2011;215:421-7.

[66] Sadeghi M, Pourmoghaddas Z, Hekmatnia A, Sanei H, Tavakoli B, Tchernof A, et al. Association of apolipoprotein B, apolipoprotein A, and the its ratio with body fat distribution. Journal of Research in Medical Sciences : The Official Journal of Isfahan University of Medical Sciences. 2013;18:326-9.

[67] Naska A, Lagiou A, Lagiou P. Dietary assessment methods in epidemiological research: current state of the art and future prospects. F1000Research. 2017;6.

Tables:

Table 1. Demographic and clinical characteristics of the total population based on the LDL and ApoB categories.

Characteristics		Group 1 (Low LDL-C, Low ApoB) (n=12,384)	Group 2 (High LDL, Low ApoB)(n=711)	Group 3 (Low LDL-C, High ApoB)(n=285)	Group 4 (High LDL-C, High ApoB)(n=885)	*p*-value
Age (Years)		48.8 ± 1.2	51.1 ± 1.9	55.6 ± 1.4	56.1 ± 1.6	<0.0001
Sex (%)	male	49.4	52.3	51.4	44.9	<0.0001
	female	50.6	47.7	49.6	55.1	
Anthropometric Parameters	BMI (kg/m²)	28.3 ± 0.1	28.1 ± 0.5	30.9 ± 0.9	30.1 ± 0.6	<0.0001
	WC (cm)	97 ± 0	97 ± 1	106 ± 2	103 ± 1	<0.0001
Insulin and Glucose Parameters	Fasting blood glucose (mg/dL)	101 ± 1	99 ± 2	118 ± 10	113 ± 4	<0.0001
	Plasma insulin (µU/mL)	13.05 ± 0.26	11.21 ± 0.64	16.77 ± 1.68	13.13 ± 0.88	<0.0001
	HOMA-IR	3.42 ± 0.08	2.76 ± 0.16	5.21 ± 0.86	3.64 ± 0.28	<0.0001
	HbA$_{1c}$ (%)	5.67 ± 0.01	5.68 ± 0.08	6.18 ± 0.30	6.20 ± 0.13	<0.0001
Liver Parameters	Alanine aminotransferase (U/L)	25 ± 1	27 ± 2	28 ± 2	30 ± 2	<0.0001
	Aspartate Aminotransferase (U/L)	26 ± 1	27 ± 2	27 ± 2	27 ± 1	<0.0001
	Gamma glutamyl transferase (U/L)	28 ± 1	37 ± 6	44 ± 7	44 ± 5	<0.0001
Macronutrients	Fat (g/day)	79 ± 1	75 ± 4	79 ± 6	69 ± 3	<0.0001
	MUFA (g/day)	29 ± 0	29 ± 2	31 ± 3	26 ± 1	<0.0001
	PUFA (g/day)	17 ± 0	15 ± 1	16 ± 1	15 ± 1	<0.0001
	SFA (g/day)	25 ± 0	25 ± 1	25 ± 2	22 ± 1	<0.0001
	Protein (g/day)	79 ± 1	78 ± 4	76 ± 6	70 ± 3	<0.0001
	Carbohydrate (g/day)	251 ± 3	238 ± 11	257 ± 14	223 ± 9	<0.0001
	Fibre (g/day)	16 ± 0	13 ± 1	14 ± 1	14 ± 1	<0.0001
	Total sugar (g/day)	114 ± 2	112 ± 7	126 ± 10	103 ± 6	<0.0001
	Energy (kcal/day)	2070 ± 21	1992 ± 88	2048 ± 116	1823 ± 69	<0.0001
Serum uric acid (mg/dL)		5.5 ± 0.0	5.7 ± 0.1	6.0 ± 0.2	5.8 ± 0.1	<0.0001
Serum CRP (mg/dL)		0.4 ± 0.0	0.3 ± 0.0	0.5 ± 0.1	0.5 ± 0.1	<0.0001
Total bilirubin (mg/dL)		0.82 ± 0.00	0.82 ± 0.02	0.80 ± 0.04	0.78 ± 0.02	<0.0001
SBP (mmHg)		122 ± 0	127 ± 2	132 ± 3	130 ± 2	<0.0001
DBP (mmHg)		68 ± 0	70 ± 1	71 ± 2	73 ± 1	<0.0001
Total cholesterol (mg/dL)		185 ± 1	251 ± 2	239 ± 3	281 ± 3	<0.0001
HDL-C (mg/DL)		54 ± 0	56 ± 1	43 ± 2	51 ± 1	<0.0001
LDL-C (mg/dL)		105 ± 1	170 ± 1	144 ± 3	192 ± 2	<0.0001
ApoB (mg/dL)		87 ± 1	119 ± 1	138 ± 1	146 ± 1	<0.0001

Triglycerides (mg/dL)	116 ± 1	117 ± 5	233 ± 13	177 ± 5	<0.0001
TG/HDL ratio	2 ± 0	2 ± 0	6 ± 0	4 ± 0	<0.0001
LDL-C/ApoB ratio	1 ± 0	1 ± 0	1 ± 0	1 ± 0	<0.0001
Non-HDL-C	131 ± 1	194 ± 1	197 ± 3	230 ± 2	<0.0001

Value expressed as a mean and SEM or percent.

Abbreviations: ApoB, apolipoprotein B; BMI, body mass index; CRP, C-reactive protein; DPB, diastolic blood pressure; HbA$_{1c}$, glycated haemoglobin; HDL-C, high-density lipoprotein cholesterol; HOMA-IR, Homeostatic Model Assessment of Insulin Resistance, LDL-C, low-density lipoprotein cholesterol, MUFA, monounsaturated fatty acids; PUFA, polyunsaturated fatty acids; SBP, systolic blood pressure; SFA, saturated fatty acids; WC, waist circumference

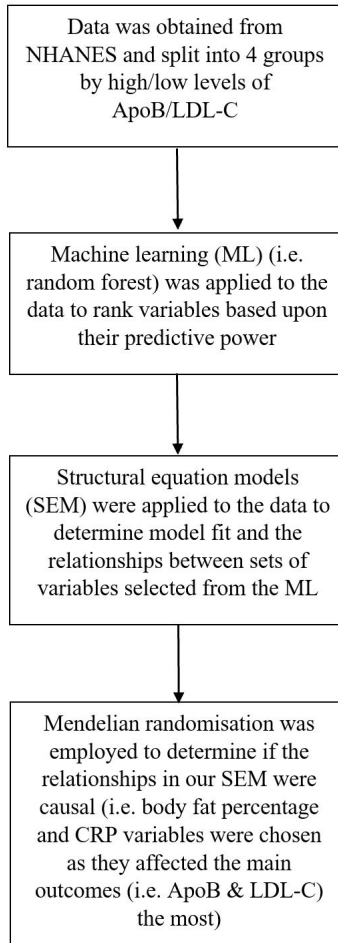**Table 2.** Results of the Mendelian Randomisation (MR) analysis for percentage body fat and CRP with LDL-C and ApoB.

Exposures		MR				Heterogeneity			Pleiotropy		
		Method	beta	SE	p	Method	Q	P-value	Intercept	SE	p
% of body fat	LDL-C	MR Egger	-0.288	0.123	0.018	MR-Egger	378.971	2.453887e-13	0.004	0.001	0.161
		WM	-0.057	0.042	0.183						
		IVW	-0.006	0.034	0.845	IVW	390.232	2.185660e-14			
		RAPS	0.016	0.035	0.643						
		MR-PRESSO	-0.001	0.031	0.971						
	ApoB	MR Egger	0.223	0.197	0.256	MR-Egger	313.226	0.0020	-0.0007	0.003	0.799
		WM	0.221	0.074	0.002						
		IVW	0.172	0.054	0.001	IVW	313.202	0.0023			
		RAPS	0.203	0.054	0.0001						
		MR-PRESSO	0.190	0.053	0.0003						
CRP	LDL-C	MR Egger	-0.097	0.093	0.308	MR-Egger	429.232	3.633e-79	0.006	0.011	0.551
		WM	-0.020	0.015	0.178						
		IVW	-0.046	0.040	0.247	IVW	437.226	3.249e-80			
		RAPS	-0.035	0.009	0.0001						
		MR-PRESSO	-0.019	0.019	0.347						
	ApoB	MR Egger	-0.108	0.096	0.270	MR-Egger	89.625	1.097e-10	0.018	0.012	0.111
		WM	-0.046	0.028	0.111						
		IVW	0.032	0.048	0.502	IVW	101.232	1.569e-12			
		RAPS	0.008	0.041	0.827						
		MR-PRESSO	-0.023	0.026	0.392						

Abbreviations: ApoB, Apolipoprotein B; beta, beta-coefficients; CRP, C-reactive protein, IVW, Inverse variance weighted; LDL-C, Low-density lipoprotein cholesterol; MR, Mendelian randomisation; SE, standard error; WM, Weighted median

1

Figure Legends:

Figure 1. Overview of study design outlining the principle methods used.

Figure 2. Structural equation model (SEM) to determine the underlying mechanism of the joined effect of ApoB and LDL-C in Group 3 (i.e. High ApoB / Low LDL-C). The diagram illustrates the SEM created to determine the underlying mechanism of joined effect of ApoB and LDL-C. The squares represent manifest nodes and arrows indicate regression coefficients which point towards an outcome of regression (standardised beta value mentioned on each arrows only for significant associations).

```
┌─────────────────────────┐
│  Data was obtained from  │
│ NHANES and split into 4  │
│   groups by high/low     │
│      levels of           │
│       ApoB/LDL-C         │
└─────────────────────────┘
            │
            ▼
┌─────────────────────────┐
│ Machine learning (ML)    │
│ (i.e. random forest) was │
│ applied to the data to   │
│ rank variables based upon│
│   their predictive power │
└─────────────────────────┘
            │
            ▼
┌─────────────────────────┐
│ Structural equation      │
│ models (SEM) were applied│
│ to the data to determine │
│ model fit and the        │
│ relationships between    │
│ sets of variables        │
│ selected from the ML     │
└─────────────────────────┘
            │
            ▼
┌─────────────────────────┐
│ Mendelian randomisation  │
│ was employed to determine│
│ if the relationships in  │
│ our SEM were causal (i.e.│
│ body fat percentage and  │
│ CRP variables were chosen│
│ as they affected the main│
│ outcomes (i.e. ApoB &    │
│ LDL-C) the most)         │
└─────────────────────────┘
```

HIGHLIGHTS

- Elevated body fat percentage is causal of increased levels of apolipoprotein B.

- Discordantly high apolipoprotein B is associated with a poor quality diet.

- Our findings demonstrate the importance of weight management in patient care.