

RESEARCH ARTICLE

Open Access



Role of circulating polyunsaturated fatty acids on cardiovascular diseases risk: analysis using Mendelian randomization and fatty acid genetic association data from over 114,000 UK Biobank participants

Maria Carolina Borges^{1,2*}, Philip C. Haycock^{1,2}, Jie Zheng^{1,2}, Gibran Hemani^{1,2}, Michael V. Holmes³, George Davey Smith^{1,2}, Aroon D. Hingorani^{4,5,6,7} and Deborah A. Lawlor^{1,2,8}

Abstract

Background: Despite early interest in the health effects of polyunsaturated fatty acids (PUFA), there is still substantial controversy and uncertainty on the evidence linking PUFA to cardiovascular diseases (CVDs). We investigated the effect of plasma concentration of omega-3 PUFA (i.e. docosahexaenoic acid (DHA) and total omega-3 PUFA) and omega-6 PUFA (i.e. linoleic acid and total omega-6 PUFA) on the risk of CVDs using Mendelian randomization.

Methods: We conducted the largest genome-wide association study (GWAS) of circulating PUFA to date including a sample of 114,999 individuals and incorporated these data in a two-sample Mendelian randomization framework to investigate the involvement of circulating PUFA on a wide range of CVDs in up to 1,153,768 individuals of European ancestry (i.e. coronary artery disease, ischemic stroke, haemorrhagic stroke, heart failure, atrial fibrillation, peripheral arterial disease, aortic aneurysm, venous thromboembolism and aortic valve stenosis).

Results: GWAS identified between 46 and 64 SNPs for the four PUFA traits, explaining 4.8–7.9% of circulating PUFA variance and with mean F statistics >100. Higher genetically predicted DHA (and total omega-3 fatty acids) concentration was related to higher risk of some cardiovascular endpoints; however, these findings did not pass our criteria for multiple testing correction and were attenuated when accounting for LDL-cholesterol through multivariable Mendelian randomization or excluding SNPs in the vicinity of the *FADS* locus. Estimates for the relation between higher genetically predicted linoleic acid (and total omega-6) concentration were inconsistent across different cardiovascular endpoints and Mendelian randomization methods. There was weak evidence of higher genetically predicted linoleic acid being related to lower risk of ischemic stroke and peripheral artery disease when accounting by LDL-cholesterol.

Conclusions: We have conducted the largest GWAS of circulating PUFA to date and the most comprehensive Mendelian randomization analyses. Overall, our Mendelian randomization findings do not support a protective role of

*Correspondence: m.c.borges@bristol.ac.uk

¹ MRC Integrative Epidemiology Unit, University of Bristol, Oakfield House, Oakfield Grove, Bristol BS8 2BN, UK

Full list of author information is available at the end of the article



© The Author(s) 2022. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

circulating PUFA concentration on the risk of CVDs. However, horizontal pleiotropy via lipoprotein-related traits could be a key source of bias in our analyses.

Keywords: Fatty acids, Cardiovascular diseases, Mendelian randomization

Background

Early interest in the cardiovascular effects of polyunsaturated fatty acids (PUFA) emerged from observational studies conducted between the 1950s and 1970s indicating that populations with a high intake of omega-3 PUFA had lower rates of mortality from cardiovascular diseases (CVDs) [1, 2]. Numerous subsequent randomized controlled trials (RCTs) indicated that dietary substitution of carbohydrates or saturated fatty acids by PUFA had a protective effect on intermediate outcomes, such as a reduction in low-density lipoprotein (LDL)-cholesterol and triglycerides [3]. However, the hypothesized cardioprotective role for omega-3 and omega-6 PUFA has been challenged by a recent series of Cochrane's systematic reviews of RCTs of dietary advice or supplementation, which suggest little to no benefit [4–6]. Overall, most RCTs on PUFA intake included in these systematic reviews were at moderate to high risk of bias and there is large uncertainty on the evidence linking PUFA to several cardiovascular outcomes [4–6]. Despite being the gold-standard study design for testing the effect of clinical interventions, in practice RCTs are often limited in statistical power, breadth of outcomes analyzed, and have high risk of bias. Therefore, integrating multiple lines of evidence is key to improve causal inference on the role of PUFAs in CVD aetiology.

Alpha-linolenic acid and linoleic acid are omega-3 and omega-6 PUFAs, respectively, that cannot be produced endogenously by humans and, therefore, need to be obtained from diet. Other omega-3 and omega-6 PUFAs can be produced endogenously through a series of elongation and desaturation reactions [7]. Circulating PUFA concentration is influenced by environmental and genetic factors [8–13], and the measurement of circulating PUFA can be used as an objective biomarker of PUFA intake, avoiding well-known biases in self-reported assessment of dietary intake. Comparing individuals genetically predisposed to higher or lower circulating PUFA can be used to probe the lifelong effect of circulating PUFA on CVD risk using Mendelian randomization [14]. Mendelian randomization uses genetic variants associated with biomarkers as instrumental variables to assess their effect on disease aetiology. This approach was developed to improve causal inference by taking advantage of unique properties of genetic variants: (i) germline genotypes remain unchanged throughout one's life, (ii) the random allocation of parental alleles at meiosis reduces

confounding by generating balanced groups and (iii) the unidirectional flow of biological information (from genotype to phenotype) avoids reverse causation [15–17]. In addition to indexing lifelong exposure to a biomarker of interest, genetic variants are subject to relatively little bias due to measurement error [18].

Previous Mendelian randomization studies have reported conflicting findings regarding the relationship between circulating PUFAs and CVDs risk [19–25]. Overall, shorter (e.g. α -linolenic acid and linoleic acid) and longer (e.g. arachidonic acid) chain PUFA have been associated with lower and higher risk of CVDs, respectively [19–25]. These seemingly contradictory findings have been largely attributed to the inclusion of genetic variants mapping to the *FADS* locus in the analyses, which contains the genes *FADS1* and *FADS2* encoding, respectively, the desaturases delta-5 (D5D) and delta-6 (D6D), key enzymes catalyzing rate-limiting steps in PUFA biosynthesis [22]. Genetic variants modulating the expression/activity of D5D and D6D will lead to changes in shorter- and longer-chain PUFAs in opposite directions, which likely explains such contrasting MR findings of lower vs higher CVD risk being associated with shorter- vs longer-chain PUFAs, respectively. In addition, these *FADS* variants are highly pleiotropic and associated with numerous non-fatty acid traits including triglycerides, low-density lipoprotein (LDL)-cholesterol and fasting glucose [22]. On the one hand, given the well-established link of *FADS1/2* with PUFA biosynthesis, single-nucleotide polymorphisms (SNPs) in the vicinity of the *FADS* locus can add to the evidence on the involvement of fatty acids in the development of cardiovascular diseases. On the other hand, the fact that SNPs nearby *FADS1/2* are highly pleiotropic and not specific for individual fatty acids or fatty acid classes complicate inferences on the causal role of circulating fatty acids on CVD risk from Mendelian randomization studies solely/pre-dominantly relying on genetic variants within this locus.

Some of these limitations in current Mendelian randomization studies could be addressed by incorporating into the analyses numerous genetic variants modulating circulating fatty acids via different pathways. Consistency of findings across variants from multiple loci would increase confidence in the results and allow the use of Mendelian randomization methods that require multiple independent variants and are more robust to violations of the method's assumptions. However,

the modest size of current genetic association studies (GWAS) on circulating fatty acids (up to ~13,000) have only allowed the discovery of a small number of genetic variants strongly and independently associated with circulating PUFA [8, 9, 11–13].

The aims of this study were (a) conducting the largest GWAS on circulating PUFA to date including a sample of 114,999 individuals and (b) using two-sample Mendelian randomization to investigate the involvement of circulating omega-3 and omega-6 fatty acids on a wide range of cardiovascular disease endpoints in up to 1,153,768 individuals of European ancestry (i.e. coronary artery disease, ischemic stroke, haemorrhagic stroke, heart failure, atrial fibrillation, peripheral arterial disease, aortic aneurysm, venous thromboembolism and aortic valve stenosis).

Methods

Data sources

Fatty acids

The UK Biobank is a population-based cohort of approximately 500,000 (~5% of those invited) people aged 40 to 69 years when recruited during 2006–2010 from several centres across the UK. A subset of approximately 20,000 were selected for repeat assessment between 2012 and 2013 [26, 27]. Details on study design, participants and quality control methods have been described previously [28, 29]. The UK Biobank received ethical approval from the Research Ethics Committee (REC reference for the UK Biobank is 11/NW/0382). This work was carried out using the UK Biobank project 30418 and 15825.

Circulating omega-3 (i.e. docosahexaenoic acid (DHA) and total omega-3) and omega-6 (i.e. linoleic acid and total omega-6) fatty acid concentration were measured using a targeted high-throughput nuclear magnetic resonance (NMR) metabolomics platform (Nightingale Health Ltd; biomarker quantification version 2020) [30]. Pre-release data from a random subset of 126,846 non-fasting plasma samples collected at baseline or first repeat assessment were made available to early access analysts. In total, 121,577 samples were retained for analyses after removing duplicates and observations not passing quality control (QC) (i.e. sample QC flag “Low protein”, biomarker QC flag “Technical error” or samples with insufficient material). This NMR platform provides simultaneous quantification of 249 metabolic measures (i.e. 165 metabolic measures and 84 derived ratios), encompassing routine lipids, lipoprotein subclass profiling (including lipid composition within 14 subclasses), fatty acid composition and various low-molecular weight metabolites such as amino acids, ketone bodies and glycolysis metabolites (Additional file 1: Table S1). Technical details and epidemiological applications of this platform have been previously reviewed [31–33]. The

mean concentration of circulating fatty acids among the UK Biobank participants was 0.23 mmol/L (SD 0.08), 0.53 mmol/L (SD 0.22), 3.41 mmol/L (SD 0.69) and 4.45 mmol/L (SD 0.68) for DHA, total omega-3, linoleic acid and total omega-6, corresponding to 2%, 4.4%, 29% and 38% of total fatty acids, respectively.

Cardiovascular disease endpoints

The outcomes of interest were (prevalent/incident) coronary artery disease, ischemic stroke, haemorrhagic stroke, heart failure, atrial fibrillation, peripheral arterial disease, aortic aneurysm, venous thromboembolism and aortic valve stenosis. Table S2 (Additional file 1) describes the data sources used for these disease endpoints.

These data sources included several large-scale genetic consortia of cardiovascular disease outcomes [34–38] targeting individuals of European ancestry only, or predominantly, including the Coronary Artery Disease Genome-Wide Replication and Meta-analysis Plus the Coronary Artery Disease Genetics Consortium (CARDIoGRAMplusC4D) (N cases = 60,801; N controls = 123,504) [34, 39], MEGASTROKE (N cases = 40,585; N controls = 406,111) [35, 40], The Heart Failure Molecular Epidemiology for Therapeutic Targets (HERMES) (N cases = 47,309; N controls = 930,014) [38, 41], an atrial fibrillation genetic association meta-analysis (N cases = 60,620; N controls = 970,216) [36, 42] and an abdominal aortic aneurysm genetic association meta-analysis (N cases = 4,972; N controls = 99,858) [37].

In addition to consortia data, we used data from two large biobanks (i.e. the UK Biobank [26, 27] and FinnGen [43]). We selected 464,708 UK Biobank participants having European genetic ancestry, as defined by an in-house k-means cluster analysis performed using the first 4 principal components provided by the UK Biobank in the statistical software environment R [29]. We used a linear mixed model (LMM) association method as implemented in BOLT-LMM (v2.3) [44] to generate genetic association data on cardiovascular endpoints among the UK Biobank participants of European genetic ancestry as described previously [29, 45]. BOLT-LMM association statistics are on the linear scale; therefore, test statistics for these binary traits (betas and their corresponding standard errors) were transformed to log odds ratios and their corresponding 95% confidence intervals on the liability scale using a Taylor transformation expansion series [44]. FinnGen is a public-private partnership project combining genotype data from Finnish biobanks and digital health record data from Finnish health registries (<https://www.finnngen.fi/en>). We used FinnGen data from release 4, which includes 176,899 participants [43]. The procedures used to generate genetic association

data for cardiovascular outcomes in FinnGen have been described previously [43].

Data analysis

Genome-wide association scan on circulating PUFA

We generated genome-wide genetic association data for circulating DHA, total omega-3, linoleic acid and total omega-6 concentration for 114,999 UK Biobank participants of European ancestry using BOLT-LMM (v2.3) [44, 46]. To model population structure in the sample, we used 143,006 directly genotyped SNPs, obtained after filtering on minor allele frequency (MAF) > 0.01; genotyping rate > 0.015; Hardy-Weinberg equilibrium p -value < 0.0001 and linkage disequilibrium (LD) pruning to an r^2 threshold of 0.1 using PLINKv2.00. Genotype array, fasting time and sex were adjusted for in the model. All fatty acid measures were standardized and normalized prior to analyses using rank-based inverse normal transformation. SNPs with MAF < 1% or imputation accuracy (INFO) < 0.8 were excluded from further analyses. The same procedures were used to generate genome-wide genetic association data for other lipid-related traits measured using the NMR metabolomics platform [i.e. total fatty acids, triglycerides, clinical LDL-cholesterol (defined to match the LDL-C levels from routine clinical chemistry) and apolipoprotein B], which were used in follow-up analyses as detailed below. The genetic association data on fatty acids and other NMR traits were deposited at the IEU Open GWAS Project [47, 48].

We used LD score regression (LDSC) to estimate genome-wide inflation in test statistics from PUFA genetic association data due to population phenomena (e.g. population stratification), as well as to approximate the genetic correlation between circulating PUFA and between PUFA and other lipid-related traits (i.e. total fatty acids, LDL-cholesterol, triglycerides and apolipoprotein B) [49, 50].

Functional mapping and annotation of PUFA genetic association results

We used FUMA GWAS (“Functional Mapping and Annotation of Genome-Wide Association Studies”), an integrative web-based platform (<http://fuma.ctglab.nl>) containing information from 18 biological data repositories and tools, to characterize SNPs according to (i) consequences to gene function, (ii) mapped genes and biological pathways and (iii) associations with other phenotypes. The pipeline used by FUMA has been described in detail elsewhere [51]. Briefly, we applied FUMA’s SNP2GENE function, which uses genome-wide genetic association data to identify independent significant SNPs and SNPs in LD with those, which are then annotated for functional consequences to gene

functions (i.e. altering expression of a gene, affecting a binding site or violating the protein structure) [52].

FUMA identifies as independent significant SNPs those associated with the trait of interest at p -value < 5×10^{-8} in the GWAS summary data and not in strong LD with each other ($R^2 < 0.6$ using EUR 1000G phase3 as reference panel). For each independent significant SNP, all known SNPs with MAF $\geq 1\%$ in strong LD ($R^2 \geq 0.6$), either present in the GWAS summary data and/or reference panel, are included for further annotation (i.e. candidate SNPs). Additionally, FUMA classifies independent lead SNPs as a subset of independent significant SNPs in weak LD ($R^2 < 0.1$) and defines genomic risk loci by merging independent significant SNPs in LD blocks and between LD blocks closely located to each other (<250 kb based on the most right and left SNPs from each LD block).

Functionally annotated SNPs are subsequently mapped to genes based on (i) physical position on the genome (positional mapping), (ii) expression quantitative trait loci (eQTL) associations (eQTL mapping) and (iii) 3D chromatin interactions (chromatin interaction mapping). In addition, independent significant SNPs and correlated SNPs are also linked to the GWAS catalog [53] to provide insights into previously reported associations of the SNPs with a variety of phenotypes. Gene-based test/gene-set analyses using MAGMA [54] are also carried out to summarize SNP associations at the gene level and associate the set of genes to biological pathways.

Selection of genetic instruments for PUFA

For our main Mendelian randomization analyses, we selected genetic variants strongly associated with circulating DHA, total omega-3 fatty acids, linoleic acid and total omega-6 fatty acids (P -value < 5×10^{-8}) as instruments for these fatty acid measures using data from 114,999 UK Biobank participants. We performed pruning to remove variants in LD ($R^2 < 0.001$ 1000G EUR population) given we used Mendelian randomization methods that assume independence between genetic instruments. For each fatty acid measure, we approximated the total R^2 and mean F statistics across selected SNPs, as previously described [55, 56], using PUFA genetic association estimates from our discovery sample (i.e. the UK Biobank) and from a replication sample (Kettunen et al. [12]) to minimize bias due to winner’s curse. The replication sample corresponds to the largest previous GWAS (median N for fatty acids 13,516) with data on circulating fatty acids measured using the same NMR metabolomics platform as in the UK Biobank participants.

Assessing the impact of genetic instruments on the fatty acid pool

To explore the specificity of the genetic instruments for individual fatty acids, we assessed their impact on the composition of circulating PUFA using data from previous GWAS in individuals of European ancestry [8, 9, 11, 12], which have modest sample size (N range 7824–13,516) but more detailed data on individual PUFAs. These GWAS have been conducted in plasma samples using different assays (i.e. gas chromatography (GC), NMR or mass spectrometry (MS)) [8, 9, 11, 12] and have expressed results in standardized molar concentration units [12], as a proportion of total fatty acids [8, 9] or in arbitrary units [11]. We used the inverse variance weighted (IVW) method to test the impact of genetically predicted DHA, total omega-3, linoleic acid and total omega-6 on individual PUFA before and after excluding SNPs within 500 kb of the *FADS* locus (chromosome 11: 61,067,097 to 62,134,826).

Multivariable regression

For comparison with Mendelian randomization results, we fitted logistic regression models to estimate the association of circulating PUFA on the risk of each cardiovascular disease endpoint using individual-level data from the UK Biobank participants, with adjustments for covariables assessed at recruitment (i.e. sex, age, body mass index (BMI), fasting time, alcohol intake, frequency, statins use and total circulating fatty acids).

Mendelian randomization

We estimated the effect of genetically predicted DHA, total omega-3, linoleic acid and total omega-6 on the risk of each cardiovascular disease endpoint using two univariable summary data Mendelian randomization methods: IVW (with multiplicative random effects) and MR-Egger [57, 58]. IVW assumes no (unbalanced) horizontal pleiotropy. There are a range of plausible circumstances where this assumption could be violated, such as in the presence of pleiotropic SNPs that influence the outcome through pathways that are not mediated by the exposure (aka horizontal pleiotropy). Because of that, we also used the MR-Egger method, which can provide valid tests of a causal effect in the presence of unbalanced horizontal pleiotropy provided that instrument strength is independent of its direct effects on the outcome (i.e. INSIDE assumption). We used Cochran's Q and the MR-Egger intercept test to explore the presence of between-SNP heterogeneity and unbalanced pleiotropy, respectively, and leave-one-out analyses to test for the presence of outlying SNPs.

Given most circulating fatty acids are carried in the circulation by lipoproteins, we anticipated that many SNPs

influencing circulating fatty acid concentration would map to genes regulating lipoprotein metabolism [12]. To mitigate bias due to lipoprotein-related traits, we used multivariable IVW to estimate direct effects of genetically predicted PUFA on the risk of cardiovascular disease endpoints after individually adjusting for total fatty acids, triglycerides, LDL-cholesterol, or apolipoprotein B, or simultaneously adjusting for both triglycerides and LDL-cholesterol. Multivariable IVW requires knowledge of the covariance between the effects of the genetic variants on each exposure, which we approximated from the phenotypic covariance matrix across exposure traits [59].

If genetic association data on cardiovascular diseases was available from two or more independent datasets (e.g. the UK Biobank and a European GWAS consortia not including the UK Biobank), we pooled these estimates using fixed-effect meta-analysis with inverse variance weights and conducted the MR analyses outlined above in both study-specific and pooled data. We used a Bonferroni correction to account for multiple testing considering the 9 different outcomes of interest (p -value = 0.05/9 outcomes = 0.006). Mendelian randomization analyses were performed using the TwoSampleMR [60] and MVMR [59] packages in R software version 3.6.2 (R Foundation for Statistical Computing).

Positive exposure control

For comparison with our Mendelian randomization analyses on circulating fatty acids, we used a positive control exposure approach, which consisted of selecting two exposures with well-established involvement in the risk of several cardiovascular diseases (i.e. LDL-cholesterol and apolipoprotein B) and conducting IVW to estimate their effect on the risk of the cardiovascular disease outcomes included in this study. Following the suggestion from a reviewer, we repeated these analyses after removing a SNP within the *FADS* locus.

Results

Genome-wide association scan on circulating PUFA

Genetic association results were available for 114,999 individuals and 12,321,876 genetic variants. FUMA identified 38, 41, 52 and 59 independent genomic regions strongly associated with circulating DHA, total omega-3, linoleic acid and total omega-6, respectively (Additional file 2: Fig. S1) [51, 52]. The Manhattan plots indicate a similar genetic association signature across PUFA, especially between DHA and total omega-3 and between linoleic acid and total omega-6 (Additional file 2: Fig. S2).

The quantile-quantile (QQ) plots did not indicate early separation of expected from observed P -values (Additional file 2: Fig. S3) and the LDSC intercepts did not indicate inflation of genome-wide test statistics due

to confounding: 1.027 (SE = 0.010), 1.037 (SE = 0.010), 1.023 (SE = 0.011) and 1.029 (SE = 0.011) for DHA, total omega-3, linoleic acid and total omega-6, respectively.

As expected, genetic correlation was high between DHA and total omega-3 (0.85 (SE = 0.02)) and between linoleic acid and total omega-6 (0.95 (SE = 0.03)) and moderate to high between omega-3 and omega-6 fatty acids (ranging from 0.39 (SE = 0.07) between DHA and linoleic acid to 0.68 (SE = 0.06) between total omega-3 and total omega-6) (Additional file 1: Table S3). Overall, genetic correlations of circulating PUFA with other NMR lipid-related traits were null to moderate for DHA (ranging from -0.01 (SE = 0.08) for triglycerides to 0.57 (SE = 0.08) for LDL-cholesterol), but moderate to high for the other PUFA measures (ranging from 0.49 (SE = 0.07) between total omega-3 and triglycerides to 0.90 (SE = 0.02) between total omega-6 and total fatty acids) (Additional file 1: Table S3).

Functional mapping and annotation of fatty acid genetic association results

Most genetic variants associated with circulating PUFA were either intronic or intergenic, suggesting that they are likely associated with PUFA concentration by modulating gene expression (Additional file 2: Fig. S1). As expected, genes mapped to PUFA-associated genetic variants included some genes with well-known role in lipoprotein metabolism, such as *CETP*, *PCSK9*, *LPL*, *LIPC*, *ANGPTL3*, *APOB*, *APOC3*, *APOA1*, *APOE* and *LDLR*, among others (Additional file 1: Tables S4A-D).

Using the full distribution of SNPs *p*-values, gene sets assigned by MAGMA were mostly related to pathways involved in lipoprotein metabolism. As an example, the top five results across PUFA measures were related to (triglyceride-rich or high-density) lipoprotein particle remodelling, apolipoprotein binding and reverse cholesterol transport (Additional file 1: Table S5A-D).

Using data from the GWAS catalog, we identified SNPs that associated with PUFA (or SNPs in strong LD with these) as correlated with numerous traits (Additional file 1: Table S6A-D). In particular, among the top ten results across PUFA were triglycerides, LDL-cholesterol, HDL-cholesterol, total cholesterol and C-reactive protein.

Selection of genetic instruments for PUFA

The number of genetic instruments selected (P -value $< 5 \times 10^{-8}$; $R^2 < 0.001$) using data from the UK Biobank participants was 46, 49, 55 and 64 SNPs for DHA, total omega-3, linoleic acid and total omega-6, respectively (listed in Additional file 1: Table S7). When combining information from SNPs present in both discovery and replication samples, the variance explained by these SNPs

in the corresponding fatty acid measure was 6.5% (DHA), 7.9% (total omega-3), 4.8% (linoleic acid) and 5.2% (total omega-6) in the UK Biobank (discovery sample) in contrast to 3.1% (DHA), 3.7% (total omega-3), 6.9% (linoleic acid) and 5.7% (total omega-6) in Kettunen et al. data [12] (replication sample). The mean *F* statistics across SNPs ranged from 109 to 201 among fatty acid measures in the UK Biobank and from 9 to 18 in the Kettunen et al. data (Additional file 1: Table S8). If excluding SNPs nearby the *FADS* locus, the variance explained by these SNPs in the corresponding fatty acid measure in the UK Biobank was 3.1% (DHA), 4.5% (total omega-3), 4.5% (linoleic acid) and 5.2% (total omega-6).

Assessing the impact of genetic instruments on the fatty acid pool

Genetic instruments predicting higher omega-3 fatty acids (i.e. DHA and total omega-3), selected from the UK Biobank participants, were associated with multiple individual fatty acids across all three independent datasets. Overall, higher genetically predicted DHA concentrations were related to higher longer chain (e.g. DHA and arachidonic acid) and lower shorter chain (e.g. α -linolenic acid and linoleic acid) PUFA (Fig. 1). After excluding SNPs nearby the *FADS* locus, genetically predicted DHA were no longer associated with most PUFA, and it became positively associated with linoleic acid (Fig. 1). Similar results were observed for total omega-3 fatty acids (Additional file 2: Fig. S4).

Genetic instruments predicting higher linoleic acid were strongly associated with increases in linoleic acid and omega-6 fatty acids, as well as DHA, in one out of the three independent datasets (Fig. 1). The exclusion of *FADS* SNPs had a modest impact on the relation between genetically predicted linoleic acid with PUFA composition (Fig. 1). Similar results were observed for total omega-6 fatty acids (Additional file 2: Fig. S4).

Multivariable regression

Characteristics of the UK Biobank participants included in multivariable regression models are summarized in Table S9 (Additional file 1).

Omega-3 PUFA (DHA and total omega-3 fatty acids)

In logistic regression models, higher circulating DHA was related to lower risk of most cardiovascular disease endpoints. The odds ratio of disease per standard unit higher DHA concentration (in models adjusted for sex, age, BMI, fasting time, alcohol intake frequency, smoking, statins use and total circulating fatty acids) was 0.91 (95% CI 0.89–0.94) for coronary artery disease, 0.90 (95% CI 0.84–0.97) for ischemic stroke, 0.92 (95% CI 0.84–1.00) for haemorrhagic stroke, 0.89 (95%

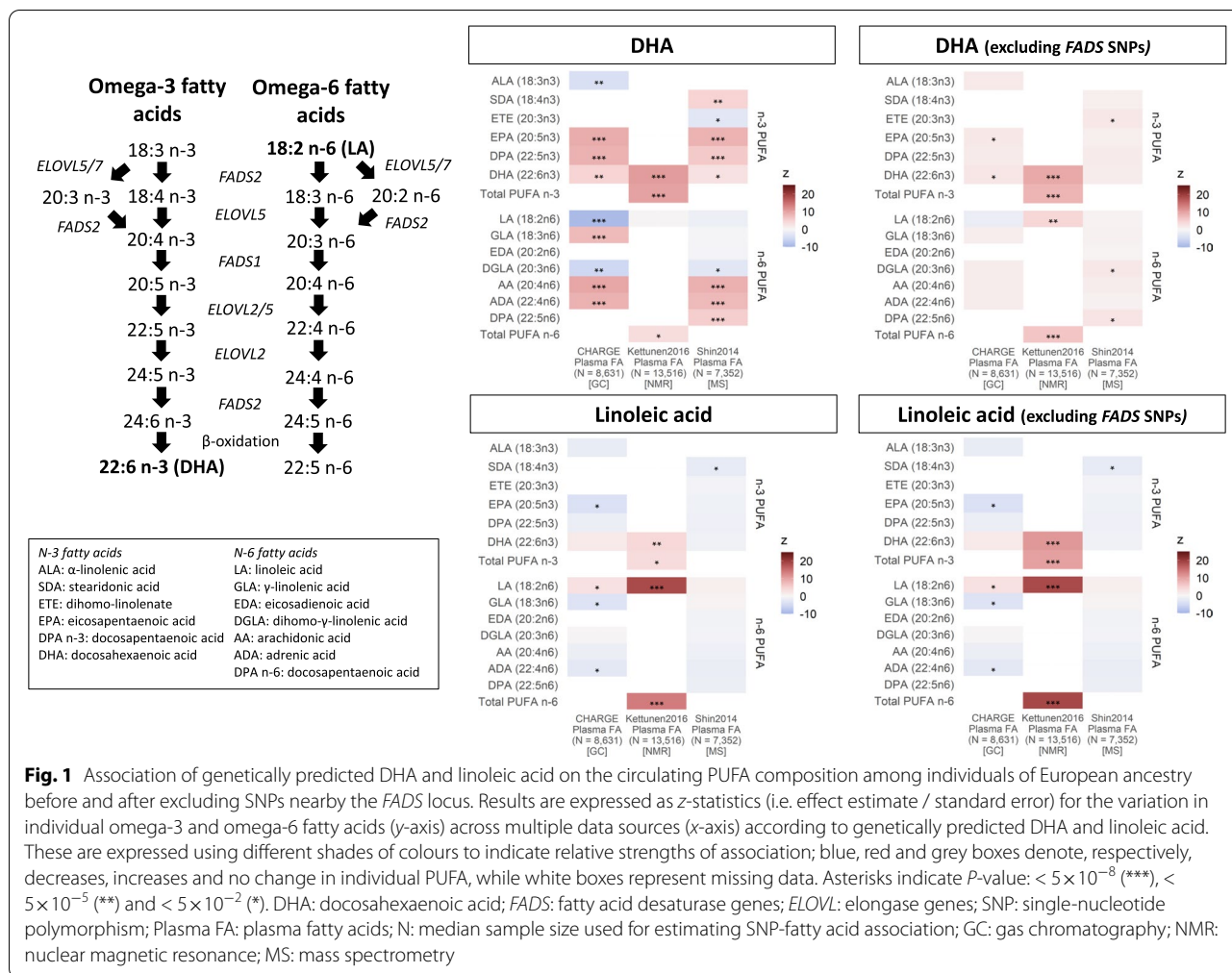


Fig. 1 Association of genetically predicted DHA and linoleic acid on the circulating PUFA composition among individuals of European ancestry before and after excluding SNPs nearby the *FADS* locus. Results are expressed as z-statistics (i.e. effect estimate / standard error) for the variation in individual omega-3 and omega-6 fatty acids (y-axis) across multiple data sources (x-axis) according to genetically predicted DHA and linoleic acid. These are expressed using different shades of colours to indicate relative strengths of association; blue, red and grey boxes denote, respectively, decreases, increases and no change in individual PUFA, while white boxes represent missing data. Asterisks indicate *P*-value: <math>P < 5 \times 10^{-8}</math> (***), <math>P < 5 \times 10^{-5}</math> (**) and <math>P < 5 \times 10^{-2}</math> (*). DHA: docosahexaenoic acid; *FADS*: fatty acid desaturase genes; *ELOVL*: elongase genes; SNP: single-nucleotide polymorphism; Plasma FA: plasma fatty acids; N: median sample size used for estimating SNP-fatty acid association; GC: gas chromatography; NMR: nuclear magnetic resonance; MS: mass spectrometry

CI 0.84–0.93) for heart failure, 0.95 (95% CI 0.92–0.98) for atrial fibrillation, 0.86 (95% CI 0.80–0.92) for peripheral artery disease, 0.87 (95% CI 0.79–0.95) for aortic aneurysm, 0.94 (95% CI 0.91–0.97) for venous thromboembolism and 1.04 (95% CI 0.95–1.14) for aortic valve stenosis. Similar results were observed for total omega-3 fatty acids (Fig. 2).

Omega-6 PUFA (linoleic acid and total omega-6 fatty acids)

Higher circulating linoleic acid was related to lower risk of most cardiovascular disease endpoints. The odds ratio of disease per standard unit higher linoleic acid concentration (in models adjusted for sex, age, BMI, fasting time, alcohol intake frequency, smoking, statins use and total circulating fatty acids) was 0.76 (95% CI 0.72–0.79) for coronary artery disease, 0.86 (95% CI 0.77–0.96) for ischemic stroke, 0.97 (95% CI 0.83–1.12) for haemorrhagic stroke, 0.75 (95% CI 0.68–0.81) for heart failure, 0.89 (95% CI 0.84–0.94) for atrial fibrillation, 0.65 (95% CI 0.58–0.72) for peripheral artery disease, 0.80 (95%

CI 0.68–0.93) for aortic aneurysm, 0.89 (95% CI 0.84–0.94) for venous thromboembolism and 0.85 (95% CI 0.74–0.99) for aortic valve stenosis. Similar results were observed for total omega-6 fatty acids (Fig. 3).

Mendelian randomization analysis

Omega-3 PUFA (DHA and total omega-3 fatty acids)

For the pooled analyses using the univariable IVW method, the odds ratio of disease per standard unit higher genetically predicted DHA concentration was 1.12 (95% CI 0.99–1.25) for coronary artery disease, 1.06 (95% CI 0.99–1.13) for ischemic stroke, 0.93 (95% CI 0.81–1.07) for haemorrhagic stroke, 1.04 (95% CI 0.98–1.11) for heart failure, 1.04 (95% CI 0.97–1.11) for atrial fibrillation, 1.14 (95% CI 1.00–1.30) for peripheral artery disease, 1.11 (95% CI 0.95–1.30) for aortic aneurysm, 1.07 (95% CI 0.96–1.19) for venous thromboembolism, 1.36 (95% CI 1.10–1.68) for aortic valve stenosis (Fig. 2). None of these results, except for aortic valve stenosis, passed our multiple testing correction threshold (*P* < 0.006).

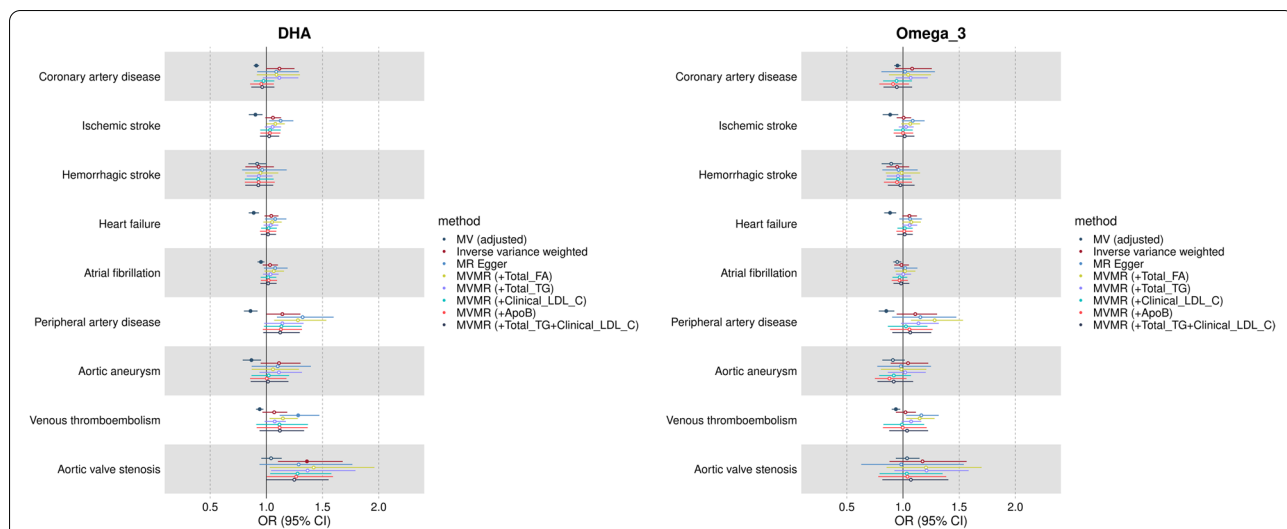


Fig. 2 Multivariable regression and Mendelian randomization results for the risk of cardiovascular diseases associated with higher genetically predicted DHA and total omega-3 fatty acids among individuals of European ancestry. Results are expressed as odds ratio (OR) of cardiovascular diseases per standard unit increase in DHA and total omega-3 fatty acids with corresponding 95% confidence interval (CI). Full symbols indicate associations at *P*-value lower than the threshold accounting multiple testing (*P* < 0.006). Multivariable regression results are adjusted by for sex, age, BMI, fasting time, alcohol intake, frequency, statins use and total circulating fatty acids. DHA: docosahexaenoic acid; MV: multivariable regression; MR: Mendelian randomization; MVMR: multivariable Mendelian randomization; FA: fatty acids; TG: triglycerides; LDL: low-density lipoprotein cholesterol; ApoB: apolipoprotein B

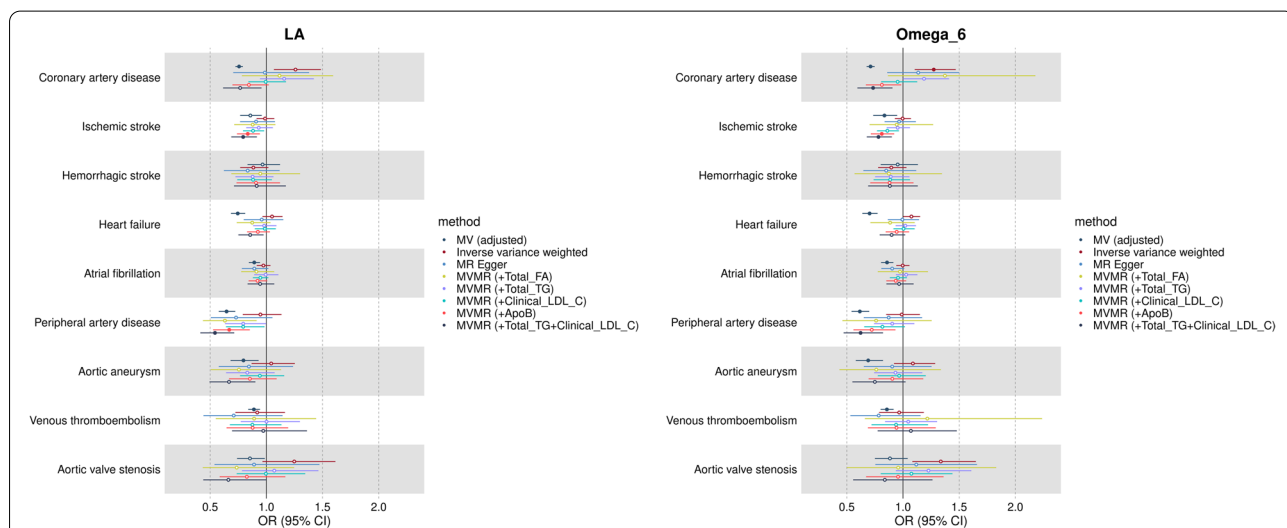


Fig. 3 Multivariable regression and Mendelian randomization results for the risk of cardiovascular diseases associated with higher genetically predicted linoleic acid and total omega-6 fatty acids among individuals of European ancestry. Results are expressed as odds ratio of cardiovascular diseases per standard unit increase in linoleic acid and total omega-6 fatty acids with corresponding 95% confidence interval (CI). Full symbols indicate associations at *P*-value lower than the threshold accounting multiple testing (*P* < 0.006). Multivariable regression results are adjusted by for sex, age, BMI, fasting time, alcohol intake, frequency, statins use and total circulating fatty acids. MV: multivariable regression; MR: Mendelian randomization; MVMR: multivariable Mendelian randomization; FA: fatty acids; TG: triglycerides; LDL: low-density lipoprotein cholesterol; ApoB: apolipoprotein B

Overall, univariable IVW point estimates for the relation between DHA and CVDs were consistent with estimates from MR-Egger, except for peripheral artery

disease and venous thromboembolism (which were slightly higher for MR-Egger), and with estimates from multivariable IVW, except for coronary artery disease

and aortic aneurism (which were fully attenuated in multivariable MR adjusting for LDL-cholesterol or apolipoprotein B) (Fig. 2).

For the vast majority of exposure-outcome combinations, there was substantial heterogeneity across individual SNP estimates as indicated by Cochran's Q statistic p -values (Q p -value range 0.04 to 2×10^{-102}), but limited evidence of unbalanced pleiotropy as assessed by the MR-Egger intercept test for most outcomes, except for peripheral artery disease (-0.015 ; p -value = 0.046) and venous thromboembolism (-0.018 ; p -value = 0.001) (Additional file 1: Table S10). Results were generally consistent across studies given imprecision in study-specific estimates (Additional file 2: Fig. S5A). Leave-one-out analyses indicated that main IVW results were driven by rs174528, a SNP nearby the *FADS* locus, since the removal of this SNP resulted in partial or complete attenuation of the IVW estimated effect for most outcomes (i.e. ischemic stroke, heart failure, atrial fibrillation, peripheral artery disease and aortic aneurism) and in reversal of the direction of the estimated effect for venous thromboembolism (Additional file 2: Fig. S6).

The conditional F statistics for DHA in multivariable IVW ranged from 76 to 112 (Additional file 1: Table S11). Results were similar between DHA and total omega-3, as shown in Fig. 2, Additional file 1: Table S10 and Additional file 2: Fig. S5A and 7.

Omega-6 PUFA (linoleic acid and total omega-6 fatty acids)

For the pooled analyses using the univariable IVW method, the odds ratio of disease per additional standard unit of genetically predicted linoleic acid concentration was 1.26 (95% CI 1.06–1.49) for coronary artery disease, 0.99 (95% CI 0.91–1.07) for ischemic stroke, 0.88 (95% CI 0.76–1.02) for haemorrhagic stroke, 1.05 (95% CI 0.96–1.14) for heart failure, 0.97 (95% CI 0.90–1.04) for atrial fibrillation, 0.95 (95% CI 0.79–1.13) for peripheral artery disease, 1.11 (95% CI 0.95–1.30) for aortic aneurysm, 0.92 (95% CI 0.72–1.17) for venous thromboembolism and 1.25 (95% CI 0.96–1.61) for aortic valve stenosis (Fig. 3). None of these results passed our multiple testing correction threshold ($P < 0.006$).

There were some inconsistencies between point estimates from univariable IVW compared to MR-Egger and multivariable IVW, though it should be noted that some estimates were very imprecise. As an example, compared to univariable IVW, estimates for the relation between circulating linoleic acid were attenuated to the null or became protective for coronary artery disease (OR 0.99; 95% CI 0.70–1.38 for MR-Egger; OR 0.77; 95% CI 0.61–0.96 for multivariable IVW simultaneously accounting for LDL-cholesterol and triglycerides) and became protective for ischemic stroke (OR

0.91; 95% CI 0.76–1.08 for MR-Egger; OR 0.79; 95% CI 0.69–0.92 for multivariable IVW accounting for triglycerides and LDL-cholesterol) and peripheral artery disease (OR 0.54; 95% CI 0.41–0.71 for multivariable IVW accounting for triglycerides and LDL-cholesterol). Overall, accounting for total fatty acids or triglycerides individually in multivariable IVW increased uncertainty (i.e. resulted in wider confidence intervals) but did not substantially change the magnitude of effect estimates as compared to univariable IVW (Fig. 3).

For vast majority of exposure-outcome combinations, there was substantial heterogeneity across individual SNP estimates as indicated by the Cochran's Q statistic p -values (Q p -value range 0.2 to 8×10^{-209}), but limited evidence of unbalanced pleiotropy as assessed by the MR-Egger intercept test [intercept ranging from 0.004 (p -value = 0.66) to 0.023 (p -value = 0.13)] (Additional file 1: Table S10). Overall, leave-one-out analyses did not indicate that the main results were substantially influenced by a single SNP with the exception of rs115478735, a SNP in the *ABO* locus, for which removal increased the magnitude and precision for the estimated effect of linoleic acid on venous thromboembolism risk (Additional file 2: Fig. S8).

The conditional F statistics for linoleic acid in multivariable IVW ranged from 52 to 128 (Additional file 1: Table S11). Results were generally consistent across studies given imprecision in study-specific estimates, except for haemorrhagic stroke and heart failure (Additional file 2: Fig. S5B), and between linoleic acid and total omega-6 (Fig. 3, Additional file 1: Table S10 and Additional file 2: Fig. S5B and S9).

Positive exposure control

Our positive control exposure analyses confirmed that higher circulating LDL-cholesterol and apolipoprotein B were related to higher risk of several cardiovascular disease outcomes. As an example, the odds ratio of disease per additional standard unit of genetically predicted LDL-cholesterol concentration was 1.55 (95% CI 1.43–1.68) for coronary artery disease, 1.13 (95% CI 1.05–1.23) for ischemic stroke, 0.91 (95% CI 0.81–1.02) for haemorrhagic stroke, 1.13 (95% CI 1.06–1.22) for heart failure, 1.04 (95% CI 0.98–1.11) for atrial fibrillation, 1.25 (95% CI 1.08–1.45) for peripheral artery disease, 1.17 (95% CI 1.02–1.35) for aortic aneurysm, 0.99 (95% CI 0.80–1.21) for venous thromboembolism and 1.35 (95% CI 1.13–1.61) for aortic valve stenosis (Additional file 2: Fig. S10). Removing a SNP within the *FADS* locus (i.e. rs102275) did not alter results from the positive control analyses (Additional file 1: Table S12).

Discussion

Using the largest-scale genetic association data available for fatty acids from over 114,000 UK Biobank participants, we identified a much larger number of genetic variants strongly and independently associated with circulating PUFA concentration compared to previous GWAS [8, 9, 11–13], which enabled us to conduct key sensitivity analyses such as MR-Egger and multivariable MR, both of which preferably require a large number of independent instruments. As expected, many of these genetic variants mapped to genes involved in lipoprotein-related metabolism. This poses a challenge to Mendelian randomization studies of fatty acids since using SNPs from a single gene region involved in fatty acid metabolism (e.g. *FADS*) is likely to be more specific but not amenable to most pleiotropy-robust Mendelian randomization methods, while using SNPs across the genome is likely to introduce non-specificity but allow the use of pleiotropy-robust methods.

When using these multiple genetic variants as instruments for Mendelian randomization, our findings did not confirm the inverse association observed in conventional multivariable regression and provided weak evidence of higher genetically predicted DHA (and total omega-3 fatty acids) concentration being related to higher risk of some cardiovascular endpoints. However, overall, Mendelian randomization findings did not pass our criteria for multiple testing correction and were attenuated when accounting for LDL-cholesterol/apolipoprotein B or excluding a SNP in the vicinity of the *FADS* locus. Mendelian randomization findings for higher genetically predicted linoleic acid (and total omega-6) concentration were inconsistent across different cardiovascular endpoints and methods and did not confirm the inverse association observed in conventional multivariable regression. There was weak evidence of higher genetically predicted linoleic acid being related to lower risk of ischemic stroke and peripheral artery disease after accounting for LDL-cholesterol/apolipoprotein B. Despite the large increase in the number of instruments in our analyses, there remains considerable imprecision in estimates for the effect of circulating fatty acids on the risk of some cardiovascular disease outcomes. As an example, IVW estimates for the relation of DHA with the risk of aortic aneurysm was 1.11 (95% CI 0.95–1.30), while IVW estimates for the relation of our positive control exposure (LDL-cholesterol) with the same outcome was 1.17 (95% CI 1.02–1.35). This indicates that we cannot confidently rule out the presence of clinically meaningful effects due to the considerable uncertainty in some results.

Previous meta-analyses of classical observational studies indicate that higher circulating long-chain omega-3

and omega-6 PUFA are either not associated or associated with lower risk of coronary artery disease and stroke [61–64]. Cochrane recently published a series of systematic reviews of RCTs with the overall conclusion that increasing omega-3, omega-6 or total PUFA intake, via supplementation or diet, has modest to no effect on CVD events or mortality [4–6]. Since then, further large-scale RCTs on long-chain omega-3 PUFA have been published and yielded conflicting results [65–67]. In addition, most Mendelian randomization studies, as well as classical observational studies and RCTs, have focussed on exploring the effect of PUFA on the risk of coronary artery disease and, to a lesser extent, ischemic stroke. The effects of these fatty acids on other types of cardiovascular endpoints, such as heart failure and atrial fibrillation, have been under explored.

Integrating multiple lines of evidence to resolve controversies in research on cardiovascular health effects of fatty acids is essential. However, directly comparing results from different study designs in this context is not straightforward. For illustration, in intervention studies, the effect of PUFA supplementation or diet intake on CVD endpoints is frequently tested over relatively short periods of time due to logistical issues [4–6] and, in some instances, may depend on the overall diet composition [3]. On the other hand, genetic proxies of circulating PUFA affect their metabolism (not intake) are assumed to have lifelong effects and have pleiotropic effects on lipoprotein-related traits.

Mendelian randomization can provide a valid test of the presence of a causal effect if genetic variants are relevant and valid instruments for the exposure of interest.

Regarding instrument relevance, we have selected independent SNPs strongly associated with circulating PUFA concentration, which explained from 4.8 to 7.9% of phenotypic variance (mean F statistics 109–201) among the UK Biobank participants (discovery sample). In addition, we replicated these associations in an independent dataset [12] using the same NMR metabolomics platform as the one used in the UK Biobank participants (median sample size 13,516), where SNPs explained 3.1 to 6.9% of phenotypic variance in circulating PUFA. This indicates that bias due to weak instruments is unlikely to be substantial in our analyses, even though bias due to winner's curse (due to using the UK Biobank to select SNPs and estimate their effect on PUFA) could affect the magnitude of effect estimates. There was little evidence that the selected SNPs impacted PUFA composition in independent datasets with more detailed data on individual PUFA, particularly after removing SNPs nearby the *FADS* locus. This might be related to differences in assays, units of analyses or statistical power.

Instrument validity requires that any effect from genetic instrument to the outcome is completely mediated by the exposure of interest. This assumption could be violated in several scenarios, such as in the presence of confounding due to population stratification or horizontal pleiotropy.

Bias by population stratification could result if heterogeneity in genetic ancestry in a given sample was related to different distributions of genetic instrument and outcome. To mitigate that, heterogeneity in genetic ancestry was accounted for when generating genetic association data for cardiovascular outcomes in all data sources by correcting for genomic inflation factor, adjusting for principal components of ancestry or using mixed linear models as detailed in Table S2 (Additional file 1). In addition, the overall consistency of findings across multiple studies provides some reassurance against findings being explained by population stratification.

Horizontal pleiotropy is one of the major threats to the validity of Mendelian randomization studies. By conducting a series of analyses for functional mapping and annotation of fatty acid genetic association results, we showed that genetic variants associated with fatty acids are strongly enriched for genes and pathways involved in lipoprotein metabolism, particularly (triglyceride-rich or high-density) lipoprotein particle remodelling, apolipoprotein binding and reverse cholesterol transport. For illustration, some PUFA SNPs mapped to genes encoding proteins targeted by lipid-lowering drugs, such as *HMGCR* (Entrez Gene 3156), *PCSK9* (Entrez Gene 255738) and *CETP* (Entrez Gene 1071) [68, 69]. Considering the pivotal role of lipoprotein metabolism in the aetiology of several cardiovascular diseases, this stresses that the assumption of no horizontal pleiotropy (or the weaker versions of this assumption by MR-Egger) in our analyses is likely implausible. We tried to mitigate that by using multivariable MR to account for total fatty acids, given the lack of specificity of the selected instruments for specific fatty acids, and for triglycerides, LDL-cholesterol or apolipoprotein B, which are key determinants of several CVDs reflecting lipoprotein metabolism. Accounting for LDL-cholesterol/apolipoprotein B revealed a potential direct protective effect of linoleic acid on the risk of ischemic stroke and peripheral artery disease, suggesting that horizontal pleiotropy via LDL-cholesterol might have masked some true underlying protective effect of linoleic acid. In addition, Mendelian randomization findings for the relation between DHA and several CVDs were attenuated when excluding SNPs in the vicinity of the *FADS* locus. On the one hand, SNPs regulating *FADS1* and *FADS2* are expected to be more credible instruments given their proximal relation with PUFA biosynthesis. However, these SNPs have been

shown to be associated with numerous fatty acids and non-fatty acid traits and may not be valid instruments for circulating DHA [22]. Results from multivariable regression were often inconsistent with results from MR IVW (e.g. multivariable regression and IVW estimates were in opposite direction for the association between PUFA and CAD risk). While the reason for that is not entirely clear, this is possibly related to pleiotropic genetic instruments since accounting for LDL-cholesterol in MR analyses attenuated those differences.

Conclusions

Overall, our Mendelian randomization findings do not support a protective role of circulating DHA, total omega-3, linoleic acid and total omega-6 concentration on the risk of CVDs, as observed in conventional analysis. However, horizontal pleiotropy via lipoprotein-related traits could be a key source of bias in our analyses despite our attempts to account for that.

Abbreviations

ALA: Alpha-linolenic acid; BMI: Body mass index; CARDIoGRAMplusC4D: Coronary Artery Disease Genome-Wide Replication and Meta-analysis Plus the Coronary Artery Disease Genetics Consortium; CVDs: Cardiovascular diseases; D5D: Delta-5 desaturase; D6D: Delta-6 desaturase; DHA: Docosahexaenoic acid; eQTL: Expression quantitative trait loci; FUMA GWAS: Functional Mapping and Annotation of Genome-Wide Association Studies; GC: Gas chromatography; GWAS: Genome-wide association study; HERMES: The Heart Failure Molecular Epidemiology for Therapeutic Targets; INSIDE assumption: Instrument Strength Independent of Direct Effects assumption; IVW method: Inverse variance weighted method; LA: Linoleic acid; LD: Linkage disequilibrium; LDL: Low-density lipoprotein; LDSC: LD score regression; LMM: Linear mixed model; MAF: Minor allele frequency; MS: Mass spectrometry; NMR: Nuclear magnetic resonance; PUFA: Polyunsaturated fatty acids; QC: Quality control; RCTs: Randomized controlled trials; SNPs: Single-nucleotide polymorphisms.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12916-022-02399-w>.

Additional file 1: Supplementary Tables S1–S11.

Additional file 2: Supplementary Figures S1–S10.

Additional file 3.

Acknowledgements

We want to acknowledge participants and investigators from FinnGen study and the multiple large-scale GWAS consortia which made summary data available, including CARDIoGRAMplusC4D, MEGASTROKE, HERMES, atrial fibrillation GWAS meta-analysis, aortic aneurysm GWAS meta-analysis and CHARGE. Data on coronary artery disease have been contributed by CARDIoGRAMplusC4D investigators and have been downloaded from www.CARDIOGRAMPLUSC4D.ORG. The MEGASTROKE project received funding from sources specified at <http://www.megastroke.org/acknowledgments.html> (a list of all MEGASTROKE authors can be found in Additional file 3). The authors are grateful to the UK Biobank for access to data to undertake this study (Project #30418 and #15825). Quality Control filtering of the UK Biobank data was conducted by R. Mitchell, G.Hemani, T. Dudding, L. Corbin, S. Harrison, L. Paternoster as described in the published protocol (doi: 10.5523/bris.1ovaa55xunp2cv8rcy88688v). The MRC IEU UK Biobank GWAS pipeline was developed by B. Elsworth,

R. Mitchell, C. Raistrick, L. Paternoster, G. Hemani, T. Gaunt (doi: 10.5523/bris.pnoat8cxo0u52p6ynfaekig). This work used the computational facilities of the Advanced Computing Research Centre, University of Bristol - <http://www.bristol.ac.uk/acrc/>.

Authors' contributions

MCB designed the study, analyzed the data and drafted the manuscript. MCB, PCH, JZ and GH contributed to the acquisition of genetic association data. PCH, JZ, GH, MVH, GDS, ADH and DAL critically revised the manuscript. All authors read and approved the final manuscript.

Funding

This research was funded by the Bristol BHF Accelerator Award (AA/18/7/34219) and the UK Medical Research Council Integrative Epidemiology Unit (MC_UU_00011/1, MC_UU_00011/4, MC_UU_00011/6 and MC_UU_00011/7). MCB was supported by a UK Medical Research Council Skills Development Fellowship (MR/P014054/1) and a Vice-Chancellor's Fellowship from the University of Bristol. PCH was supported by Cancer Research UK (C52724/A20138 & C18281/A29019). JZ is funded by a Vice-Chancellor Fellowship from the University of Bristol and is supported by the Academy of Medical Sciences (AMS) Springboard Award, the Wellcome Trust, the Government Department of Business, Energy and Industrial Strategy (BEIS), the British Heart Foundation and Diabetes UK (SBF006\1117). GH was funded by the Wellcome Trust and the Royal Society [208806/Z/17/Z]. MVH works in a unit that receives funding from the UK Medical Research Council and is supported by a British Heart Foundation Intermediate Clinical Research Fellowship (FS/18/23/33512) and the National Institute for Health Research Oxford Biomedical Research Centre. GDS is supported by the National Institute for Health Research (NIHR) Biomedical Research Centre based at University Hospitals Bristol NHS Foundation and the University of Bristol. ADH is a National Institute of Health Research Senior Investigator. DAL is BHF Chair of Cardiovascular Science and Clinical Epidemiology (CH/F/20/90003) and a National Institute of Health Research Senior Investigator (NF-0616-10102). No funder influenced the study design, analyses or interpretation of results. The views expressed in the paper are those of the authors and not necessarily of the listed funders.

Availability of data and materials

The genetic association data on fatty acids and other NMR metabolic traits generated by this study were deposited at the IEU Open GWAS Project [47, 48]. Genetic association data from GWAS meta-analyses and FinnGen for cardiovascular endpoints from contributing studies is publicly available as detailed in Additional file 1: Table S2 [39–43]. Researchers can apply for access to the UK Biobank data via the Access management System (AMS) (<https://www.ukbiobank.ac.uk/enable-your-research/apply-for-access>). The code for the MR analyses is available at [https://github.com/mcarolborges/NMR_PUFA.git].

Declarations

Ethics approval and consent to participate

The UK Biobank received ethical approval from the Research Ethics Committee (REC reference for the UK Biobank is 11/NW/0382).

Consent for publication

Not applicable.

Competing interests

DAL is an Editorial Advisor for BMC Medicine and acknowledges support from Roche Diagnostics and Medtronic Ltd for research unrelated to that presented here. All other authors declare they have no conflict of interest, financial or otherwise.

Author details

¹MRC Integrative Epidemiology Unit, University of Bristol, Oakfield House, Oakfield Grove, Bristol BS8 2BN, UK. ²Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK. ³MRC Population Health Research Unit (MRC PHRU), Nuffield Department of Population Health, University of Oxford, Oxford, UK. ⁴Institute of Cardiovascular Science, Faculty of Population Health Sciences, University College London, London, UK. ⁵UCL BHF Research Accelerator, London, UK. ⁶Health Data Research UK, Institute of Health Informatics,

University College London, London, UK. ⁷UCL NIHR Biomedical Research Centre, London, UK. ⁸NIHR Bristol Biomedical Research Centre, Bristol, UK.

Received: 29 September 2021 Accepted: 9 May 2022

Published online: 13 June 2022

References

- Bang HO, Dyerberg J, Sinclair HM. The composition of the Eskimo food in north western Greenland. *Am J Clin Nutr*. 1980;33(12):2657–61.
- Kagawa Y, Nishizawa M, Suzuki M, Miyatake T, Hamamoto T, Goto K, et al. Eicosapolyenoic acids of serum lipids of Japanese islanders with low incidence of cardiovascular diseases. *J Nutr Sci Vitaminol (Tokyo)*. 1982;28(4):441–53.
- Mensink RP, Zock PL, Kester AD, Katan MB. Effects of dietary fatty acids and carbohydrates on the ratio of serum total to HDL cholesterol and on serum lipids and apolipoproteins: a meta-analysis of 60 controlled trials. *Am J Clin Nutr*. 2003;77(5):1146–55.
- Abdelhamid AS, Martin N, Bridges C, Brainard JS, Wang X, Brown TJ, et al. Polyunsaturated fatty acids for the primary and secondary prevention of cardiovascular disease. *Cochrane Database Syst Rev*. 2018;11(11):CD012345.
- Hooper L, Al-Khudairy L, Abdelhamid AS, Rees K, Brainard JS, Brown TJ, et al. Omega-6 fats for the primary and secondary prevention of cardiovascular disease. *Cochrane Database Syst Rev*. 2018;11(11):CD011094.
- Abdelhamid AS, Brown TJ, Brainard JS, Biswas P, Thorpe GC, Moore HJ, et al. Omega-3 fatty acids for the primary and secondary prevention of cardiovascular disease. *Cochrane Database Syst Rev*. 2020;3(3):CD003177.
- Jakobsson A, Westerberg R, Jakobsson A. Fatty acid elongases in mammals: their regulation and roles in metabolism. *Prog Lipid Res*. 2006;45(3):237–49.
- Lemaitre RN, Tanaka T, Tang W, Manichaikul A, Foy M, Kabagambe EK, et al. Genetic loci associated with plasma phospholipid n-3 fatty acids: a meta-analysis of genome-wide association studies from the CHARGE Consortium. *PLoS Genet*. 2011;7(7):e1002193.
- Guan W, Steffen BT, Lemaitre RN, Wu JH, Tanaka T, Manichaikul A, et al. Genome-wide association study of plasma N6 polyunsaturated fatty acids within the cohorts for heart and aging research in genomic epidemiology consortium. *Circ Cardiovasc Genet*. 2014;7(3):321–31.
- Wu JH, Lemaitre RN, Manichaikul A, Guan W, Tanaka T, Foy M, et al. Genome-wide association study identifies novel loci associated with concentrations of four plasma phospholipid fatty acids in the de novo lipogenesis pathway: results from the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) consortium. *Circ Cardiovasc Genet*. 2013;6(2):171–83.
- Shin SY, Fauman EB, Petersen AK, Krumsiek J, Santos R, Huang J, et al. An atlas of genetic influences on human blood metabolites. *Nat Genet*. 2014;46(6):543–50.
- Kettunen J, Demirkan A, Wurtz P, Draisma HH, 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.
- Tintle NL, Pottala JV, Lacey S, Ramachandran V, Westra J, Rogers A, et al. A genome-wide association study of saturated, mono- and polyunsaturated red blood cell fatty acids in the Framingham Heart Offspring Study. *Prostaglandins Leukot Essent Fatty Acids*. 2015;94:65–72.
- Smith GD, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *Int J Epidemiol*. 2003;32(1):1–22.
- Ebrahim S, Davey Smith G. Mendelian randomization: can genetic epidemiology help redress the failures of observational epidemiology? *Hum Genet*. 2008;123(1):15–33.
- Hingorani A, Humphries S. Nature's randomised trials. *Lancet*. 2005;366(9501):1906–8.
- Smith GD, Lawlor DA, Harbord R, Timpson N, Day I, Ebrahim S. Clustered environments and randomized genes: a fundamental distinction between conventional and genetic epidemiology. *PLoS Med*. 2007;4(12):e352.
- Davey Smith G, Hemani G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Hum Mol Genet*. 2014;23(R1):R89–98.

19. Jäger S, Cuadrat R, Hoffmann P, Wittenbecher C, Schulze MB. Desaturase activity and the risk of type 2 diabetes and coronary artery disease: a Mendelian randomization study. *Nutrients*. 2020;12(8):2261.
20. Liao LZ, Li WD, Liu Y, Li JP, Zhuang XD, Liao XX. Exploring the causal pathway from omega-6 levels to coronary heart disease: a network Mendelian randomization study. *Nutr Metab Cardiovasc Dis*. 2020;30(2):233–40.
21. Yuan T, Si S, Li Y, Li W, Chen X, Liu C, et al. Roles for circulating polyunsaturated fatty acids in ischemic stroke and modifiable factors: a Mendelian randomization study. *Nutr J*. 2020;19(1):70.
22. Yuan S, Bäck M, Bruzelius M, Mason AM, Burgess S, Larsson S. Plasma phospholipid fatty acids, FADS1 and risk of 15 cardiovascular diseases: a Mendelian randomisation study. *Nutrients*. 2019;11(12):3001.
23. Zhang T, Zhao JV, Schooling CM. The associations of plasma phospholipid arachidonic acid with cardiovascular diseases: a Mendelian randomization study. *EBioMedicine*. 2021;63:103189.
24. Chen HY, Cairns BJ, Small AM, Burr HA, Ambikumar A, Martinsson A, et al. Association of FADS1/2 locus variants and polyunsaturated fatty acids with aortic stenosis. *JAMA Cardiol*. 2020;5(6):694–702.
25. Yuan S, Larsson SC. Plasma phospholipid fatty acids and risk of atrial fibrillation: a Mendelian randomization study. *Nutrients*. 2019;11(7):1651.
26. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med*. 2015;12(3):e1001779.
27. Allen NE, Sudlow C, Peakman T, Collins R, Biobank UK. UK biobank data: come and get it. *Sci Transl Med*. 2014;6(224):224ed224.
28. Collins R. What makes UK Biobank special? *Lancet*. 2012;379(9822):1173–4.
29. Mitchell R, Hemani G, Dudding T, Corbin L, Harrison S, Paternoster L. UK Biobank Genetic Data: MRC-IEU Quality Control, version 2 - Datasets - data.bris. data.bris; 2018.
30. Julkunen H, Cichonska A, Slagboom PE, Wurtz P. Nightingale Health UKBI: Metabolic biomarker profiling for identification of susceptibility to severe pneumonia and COVID-19 in the general population. *Elife*. 2021;10:e63033.
31. Soininen P, Kangas AJ, Wurtz P, Suna T, Ala-Korpela M. Quantitative serum nuclear magnetic resonance metabolomics in cardiovascular epidemiology and genetics. *Circ Cardiovasc Genet*. 2015;8(1):192–206.
32. Soininen P, Kangas AJ, Wurtz P, Tukiainen T, Tykkynen T, Laatikainen R, et al. High-throughput serum NMR metabolomics for cost-effective holistic studies on systemic metabolism. *Analyst*. 2009;134(9):1781–5.
33. Würtz P, Kangas AJ, Soininen P, Lawlor DA, Davey Smith G, Ala-Korpela M. Quantitative serum nuclear magnetic resonance metabolomics in large-scale epidemiology: a primer on -Omic technologies. *Am J Epidemiol*. 2017;186(9):1084–96.
34. Nikpay M, Goel A, Won HH, Hall LM, Willenborg C, Kanoni S, et al. A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease. *Nat Genet*. 2015;47(10):1121–30.
35. Malik R, Chauhan G, Traylor M, Sargurupremraj M, Okada Y, Mishra A, et al. Multi-ancestry genome-wide association study of 520,000 subjects identifies 32 loci associated with stroke and stroke subtypes. *Nat Genet*. 2018;50(4):524–37.
36. Nielsen JB, Thorolfsdottir RB, Fritsche LG, Zhou W, Skov MW, Graham SE, et al. Biobank-driven genomic discovery yields new insight into atrial fibrillation biology. *Nat Genet*. 2018;50(9):1234–9.
37. Jones GT, Tromp G, Kuivaniemi H, Gretarsdottir S, Baas AF, Giusti B, et al. Meta-analysis of genome-wide association studies for abdominal aortic aneurysm identifies four new disease-specific risk loci. *Circ Res*. 2017;120(2):341–53.
38. Shah S, Henry A, Roselli C, Lin H, Sveinbjornsson G, Fatemifar G, et al. Genome-wide association and Mendelian randomisation analysis provide insights into the pathogenesis of heart failure. *Nat Commun*. 2020;11(1):163.
39. CARDIoGRAMplusC4D 1000 Genomes-based GWAS summary data (2015) [<http://www.cardiogramplusc4d.org/data-downloads/>]
40. MEGASTROKE GWAS summary data (2018) [<http://www.megastroke.org/>]
41. HERMES Heart Failure GWAS summary data (2020) [<https://kp4cd.org/datasets/mi>]
42. Atrial Fibrillation GWAS summary data (2018) [<http://csg.sph.umich.edu/willer/public/afib2018/>]
43. FinnGen. Documentation of R4 release [<https://finngen.gitbook.io/documentation/>]
44. Loh PR, Kichaev G, Gazal S, Schoech AP, Price AL. Mixed-model association for biobank-scale datasets. *Nat Genet*. 2018;50(7):906–8.
45. Mitchell REB, Raistrick CA, Paternoster L, Hemani G, Gaunt TR. MRC IEU UK Biobank GWAS pipeline version 2; 2019.
46. Loh PR, Tucker G, Bulik-Sullivan BK, Vilhjalmsson BJ, Finucane HK, Salem RM, et al. Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nat Genet*. 2015;47(3):284–90.
47. Elsworth B, Lyon M, Alexander T, Liu Y, Matthews P, Hallett J, et al. The MRC IEU OpenGWAS data infrastructure. *bioRxiv*. 2020;2020:2008.2010.244293.
48. MRC Integrative Epidemiology Unit. IEU Open GWAS Project [<https://gwas.mrcieu.ac.uk/>]
49. Bulik-Sullivan BK, Loh PR, Finucane HK, Ripke S, Yang J, Schizophrenia Working Group of the Psychiatric Genomics C, et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet*. 2015;47(3):291–5.
50. Bulik-Sullivan B, Finucane HK, Anttila V, Gusev A, Day FR, Loh PR, et al. An atlas of genetic correlations across human diseases and traits. *Nat Genet*. 2015;47(11):1236–41.
51. Watanabe K, Taskesen E, van Bochoven A, Posthuma D. Functional mapping and annotation of genetic associations with FUMA. *Nat Commun*. 2017;8(1):1826.
52. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res*. 2010;38(16):e164.
53. Buniello A, MacArthur JAL, Cerezo M, Harris LW, Hayhurst J, Malagone C, et al. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res*. 2019;47(D1):D1005–12.
54. de Leeuw CA, Mooij JM, Heskes T, Posthuma D. MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput Biol*. 2015;11(4):e1004219.
55. Bowden J, Del Greco MF, Minelli C, Davey Smith G, Sheehan NA, Thompson JR. Assessing the suitability of summary data for two-sample Mendelian randomization analyses using MR-Egger regression: the role of the I² statistic. *Int J Epidemiol*. 2016;45(6):1961–74.
56. Burgess S, Davies NM, Thompson SG. Bias due to participant overlap in two-sample Mendelian randomization. *Genet Epidemiol*. 2016;40(7):597–608.
57. 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(2):512–25.
58. Bowden J, Del Greco MF, Minelli C, Davey Smith G, Sheehan N, Thompson J. A framework for the investigation of pleiotropy in two-sample summary data Mendelian randomization. *Stat Med*. 2017;36(11):1783–802.
59. Sanderson E, Davey Smith G, Windmeijer F, Bowden J. An examination of multivariable Mendelian randomization in the single-sample and two-sample summary data settings. *Int J Epidemiol*. 2019;48(3):713–27.
60. Hemani G, Zheng J, Elsworth B, Wade KH, Haberland V, Baird D, et al. The MR-Base platform supports systematic causal inference across the human phenome. *Elife*. 2018;7:e34408.
61. Borges MC, Schmidt AF, Jefferis B, Wannamethee SG, Lawlor DA, Kivimaki M, et al. Circulating fatty acids and risk of coronary heart disease and stroke: individual participant data meta-analysis in up to 16 126 participants. *J Am Heart Assoc*. 2020;9(5):e013131.
62. Chowdhury R, Warnakula S, Kunutsor S, Crowe F, Ward HA, Johnson L, et al. Association of dietary, circulating, and supplement fatty acids with coronary risk: a systematic review and meta-analysis. *Ann Intern Med*. 2014;160(6):398–406.
63. Chowdhury R, Stevens S, Gorman D, Pan A, Warnakula S, Chowdhury S, et al. Association between fish consumption, long chain omega 3 fatty acids, and risk of cerebrovascular disease: systematic review and meta-analysis. *BMJ*. 2012;345:e6698.
64. Iso H, Sato S, Umemura U, Kudo M, Koike K, Kitamura A, et al. Linoleic acid, other fatty acids, and the risk of stroke. *Stroke*. 2002;33(8):2086–93.
65. Bhatt DL, Steg PG, Miller M, Brinton EA, Jacobson TA, Ketchum SB, et al. Cardiovascular risk reduction with icosapent ethyl for hypertriglyceridemia. *N Engl J Med*. 2019;380(1):11–22.
66. Manson JE, Cook NR, Lee IM, Christen W, Bassuk SS, Mora S, et al. Marine n-3 fatty acids and prevention of cardiovascular disease and cancer. *N Engl J Med*. 2019;380(1):23–32.

67. Nicholls SJ, Lincoff AM, Garcia M, Bash D, Ballantyne CM, Barter PJ, et al. Effect of high-dose omega-3 fatty acids vs corn oil on major adverse cardiovascular events in patients at high cardiovascular risk: the STRENGTH randomized clinical trial. *JAMA*. 2020;324(22):2268–80.
68. Ference BA, Kastelein JJP, Ginsberg HN, Chapman MJ, Nicholls SJ, Ray KK, et al. Association of genetic variants related to CETP inhibitors and statins with lipoprotein levels and cardiovascular risk. *JAMA*. 2017;318(10):947–56.
69. Ference BA, Robinson JG, Brook RD, Catapano AL, Chapman MJ, Neff DR, et al. Variation in PCSK9 and HMGCR and risk of cardiovascular disease and diabetes. *N Engl J Med*. 2016;375(22):2144–53.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

