Investigating the Relationship between Polyunsaturated Fatty Acids and Inflammation using Cohort and Two-sample Mendelian Randomization Analyses - Supplementary Material

Contents

[1. eMethods 1](#_Toc151622443)

[1.1. ALSPAC recruitment details 2](#_Toc151622444)

[1.2. Covariates 2](#_Toc151622445)

[1.3. Multiple Imputation 3](#_Toc151622446)

[1.4. Mendelian randomization 3](#_Toc151622447)

[1.5. Generating fatty acid genetic instruments 4](#_Toc151622448)

[1.6. Instrument Selection for Univariable MR 4](#_Toc151622449)

[1.7. MR-Lap analysis to investigate bias from sample overlap, weak instrument and winner’s curse 5](#_Toc151622450)

[1.8. Instrument selection for specific PUFA genes 5](#_Toc151622451)

[1.9. MR-Clust analysis to explore heterogeneity in MR instruments 6](#_Toc151622452)

[1.10. Multivariable mendelian randomization analysis 7](#_Toc151622453)

[2. eResults 7](#_Toc151622454)

[2.1. Instrument validity 7](#_Toc151622455)

[2.2. Results from investigating biases impacting MR using MR-Lap 8](#_Toc151622456)

[2.3. Potential causal effect of DHA levels on inflammatory markers 8](#_Toc151622457)

[2.3.1. Effect of DHA on outcome CRP levels 8](#_Toc151622458)

[2.3.2. Effect of DHA on outcome GlycA levels 8](#_Toc151622459)

[2.4. Potential causal effect of LA levels on inflammatory markers 9](#_Toc151622460)

[2.4.1. Effect of LA on outcome CRP levels 9](#_Toc151622461)

[2.4.2. Effect of LA on outcome GlycA levels 9](#_Toc151622462)

[2.5. Potential causal effect of n-3 levels on inflammatory markers 9](#_Toc151622463)

[2.5.1. Effect of total n-3 PUFAs on outcome CRP levels 9](#_Toc151622464)

[2.5.2. Effect of total n-3 PUFAs on outcome GlycA levels 9](#_Toc151622465)

[2.6. Potential causal effect of n-6 levels on inflammatory markers 10](#_Toc151622466)

[2.6.1. Effect of total n-6 PUFAs on outcome CRP levels 10](#_Toc151622467)

[2.6.2. Effect of total n-6 PUFAs on outcome GlycA levels 10](#_Toc151622468)

[2.7. Results from MR analysis focusing on specific PUFA genes 10](#_Toc151622469)

[2.8. Results from MR-Clust analysis to explore heterogeneity in MR instruments 11](#_Toc151622470)

[2.9. MVMR results 11](#_Toc151622471)

[3. eReferences 13](#_Toc151622472)

# eMethods

## ALSPAC recruitment details

Please note that the ALSPAC website contains details of all the data that is available through a fully searchable data dictionary and variable search tool http://www.bristol.ac.uk/alspac/researchers/our-data/. Ethical approval for the study was obtained from the ALSPAC Ethics and Law Committee and the Local Research Ethics Committee and can be found here: <http://www.bristol.ac.uk/alspac/researchers/research-ethics/>. Informed consent for the use of data collected via questionnaires and clinics was obtained from participants following the recommendations of the ALSPAC Ethics and Law Committee at the time. Consent for biological samples has been collected in accordance with the Human Tissue Act (2004). Study data were collected and managed using REDCap electronic data capture tools hosted at the University of Bristol. REDCap (Research Electronic Data Capture) is a secure, web-based software platform designed to support data capture for research studies5.

## Covariates

Childhood socio-economic position (SEP) was indexed using maternal self-reported highest education qualification at 8-42 weeks gestation (dichotomised according to presence/absence of a degree level qualification); and the highest social occupation of either parent (dichotomised according to manual/non-manual occupation). Maternal and paternal smoking pattern during pregnancy (categorised as no smoking: no reports of smoking in any trimester and, any smoking: reports of smoking in any trimester)) was used as a proxy for general pregnancy health. Participants’ sex was recorded and age (in months), participants’ smoking status and drinking status (categorised as none or ever) were reported at the assessment clinic. These covariates were selected *a priori* based on known or plausible relationships with both systemic inflammation and polyunsaturated fatty acids (PUFAs).

## Multiple Imputation

We created 16 datasets with 160 iterations which aimed to balance computational efficiency while ensuring consistency of results6. The main and auxiliary variables used to inform the imputation are listed in eTable 21. Imputation was run in Stata version 17.0 (Stata Corp, College Station, TX) and estimates were then combined through Rubin's rules. We compared imputed vs non-imputed characteristics to check the imputation model.

## Mendelian randomization

Mendelian Randomisation (MR)1, 2 is an instrumental variable analysis that uses the random assortment of genetic variants from parents to offspring to investigate the causal effect of an exposure on an outcome3. MR uses genetic variants as instrumental variables to estimate causal effects which, if key assumptions are met, can overcome issues of reverse causation and unmeasured confounding. MR is bound by three assumptions (i) the relevance assumption, which describes how the genetic variants used to instrument the exposure of interest are statistically strongly associated with the exposure of interest and relevant to the population to which inference is being made; (ii) the independence assumption, which suggests that there is no confounding of the genetic instrument-outcome association; and (iii) the exclusion-restriction assumption, which suggests that the effect of the genetic instrument on the outcome is only via the exposure of interest. As such, conventional multivariable regression analysis and MR have differing and unrelated sources of bias and can be used to triangulate evidence where, when results are concordant, they can be used to support causal inference 4.

## Generating fatty acid genetic instruments

Genetic variant (single nucleotide polymorphism [SNP]) instrument sets were generated for each fatty acid phenotype of interest, namely circulating levels of docosahexaenoic acid (DHA), linoleic acid (LA), total n-3 PUFAs and total n-6 PUFAs, using summary level data from European population genome-wide association studies (GWAS) which used data from UK Biobank (UKB; n=114,999). For the outcome variables, we used SNPs identified in summary-level data from European population GWAS of the inflammatory biomarkers GlycA7, CRP8 and IL-69. The GlycA GWAS used data from UKB (n=114,999), the serum CRP GWAS used data from the Cohort for Heart and Aging Research in Genomic Epidemiology (CHARGE; n=575,531)10 and the IL-6 GWAS used data from a meta-analysis of 26 cohorts with a total of 52,654 individuals9 . Ethical approval was obtained in all original studies.

## Instrument Selection for Univariable MR

Using the TwosampleMR R package, fatty acid SNP instrument information (e.g. SNP alleles, effect allele frequency etc.) was harmonised with the outcome inflammatory biomarker information and therefore the SNP-exposure and SNP-outcome effects corresponded to the same allele. During harmonization, palindromic SNPs with intermediate allele frequencies of above 0.42 and below 0.58 were considered strand ambiguous and removed and SNPs that were unavailable in the outcome datasets were replaced by proxy SNPs at R2>0.80. Further, SNPs that had a minor allele frequency of 0.01 or less were excluded and we also ensured that the SNPs were independent by clumping the harmonized SNPs for linkage disequilibrium (LD, r2=0.001). Not all SNPs were available in every GWAS which lead to a different number of available SNPs for each analysis.

## MR-Lap analysis to investigate bias from sample overlap, weak instrument and winner’s curse

MR-Lap is a novel method that assesses potential bias in MR effect estimates due to sample overlap, weak instrument bias and winner’s curse11. If the MR-Lap corrected effect does not significantly differ from the observed effect, then the IVW-MR estimate can be used. If there is a difference between the two estimates, it suggests that the biases are having an effect on the effect estimate and therefore, the corrected effect would be preferred 11. For this analysis, we used the same PUFA and inflammatory biomarker GWAS’ as used in the main analysis. Instrumental variables (IVs) were pruned to the distance threshold of 10000 Kb with an LD threshold of 0.01.

## Instrument selection for specific PUFA genes

Instrument variables which fell in/near (+/- 500kb) to the genes that encode the key fatty acid desaturase enzymes (FADS gene cluster of *FADS1*, *FADS2* and *FADS3*) (chromosome 11: 61 560 452–61 659 523) and elongase enzymes *ELOVL2* gene locus *(*chromosome 6: 10 980 992–11 044 624), were extracted from the DHA GWAS and the LA GWAS summary statistics.

A total of 2026 independent SNPs reached genome-wide significance (P< 5∙0 x10-8) within the *FADS* gene cluster of the DHA GWAS and the LA GWAS. However, after clumping using a European population reference to remove SNPs in LD, 11 SNPs remained from the DHA GWAS and 7 genome-wide significant SNPs remained available from the LA GWAS.

A total of 5399 SNPs reached genome-wide significance within the *ELOVL2* gene of the DHA GWAS and the LA GWAS. However, after clumping using a European population reference to remove SNPs in LD, 9 genome-wide significant SNPs remained available from the DHA GWAS and 17 genome-wide significant SNPs remained available from the LA GWAS .

SNPs were harmonized, aligning the genetic association for exposure and outcome on the effect allele using the effect allele frequency. As in the primary analysis, palindromic SNPs with intermediate allele frequencies of above 0.42 and below 0.58 were considered strand ambiguous and removed, and SNPs that were unavailable in the outcome datasets were replaced by proxy SNPs at R2>0.80. Further, SNPs that had a minor allele frequency of 0.01 or less were excluded and we ensured that the SNPs were independent by clumping the harmonized SNPs for LD (r2=0.001). Steiger filtering was applied. The number of available SNPs for each analysis are presented in eTable 22.

Following harmonization, there were 1-11 SNPs available to investigate the effect of the *FADS* gene cluster from the DHA GWAS on CRP, GlycA and IL-6 levels. There were 6 and 7 SNPs respectively available to investigate the effect of the *FADS* gene cluster from the LA GWAS on CRP and GlycA levels, but no *FADS* gene cluster SNPs from the LA GWAS were available for the analysis with IL-6 as the outcome.

When using SNPs identified to be within the *ELOVL2* gene, following harmonization, there were 2-7 SNPs available from the DHA GWAS to investigate the effect of the *ELOVL2* gene on CRP, GlycA and IL-6 levels. There were 1-14 SNPs available to investigate the effect of the *ELOVL2* gene from the LA GWAS on CRP, GlycA and IL-6 levels. Harmonized SNP information for each relationship is presented in eTable 23 and eTable 24.

## MR-Clust analysis to explore heterogeneity in MR instruments

Using the author’s guidelines we opted to use the conservative method, whereby only variants were assigned to a cluster if the conditional probability of cluster assignment was > 0.89 and clusters were only recognised if they contained at least four variants. Variants that did not satisfy these criteria and did not fall into a null cluster were grouped into a “junk” cluster. The null cluster was a collection of variants that did not have an effect on the outcome.

## Multivariable mendelian randomization analysis

Linkage disequilibrium score regression (LDSC) estimated the genetic correlation between total n-3 and total n-6 (0.334) which confirmed our assumption that they may act as a pleiotropic pathway. We used the same GWAS data as applied in the univariable MR.

The same thresholds for genome-wide significance, LD and palindromic SNPs were applied as in the univariable MR. The n-3 SNPs and n-6 SNPs were combined giving a final exposure instrument of 158 SNPs. The exposure instrument was then harmonised with the outcome data (GlycA/CRP/IL-6) and after harmonization there were 31 SNPs for the exposures total n-3 PUFAs and total n-6 PUFAs on the outcome CRP, 78 SNPs for the exposures total n-3 PUFAs and total n-6 PUFAs on the outcome GlycA and 9 SNPs for the exposures total n-3 PUFAs and total n-6 PUFAs on IL-6.

# eResults

## Instrument validity

Within the univariable analysis, SNP-level F statistics were all >10, indicating that IVW estimates were likely not subject to weak instrument bias. The total variance explained (R2) by each SNP in the univariable MR was between 0.03% and 4.03% for DHA, 0.03% and 0.55% for LA, 0.02% and 5.21% for total n-3 PUFAs, and 0.02% and 0.62% for total n-6 PUFAs. Genetic instruments, sample sizes and instrument F-statistics for each GWAS are presented in eTable 25. All results are presented after Steiger filtering.

## Results from investigating biases impacting MR using MR-Lap

Although recently published research has suggested that sample overlap does not incur as much bias as previously thought11, we used MR-lap to investigate whether sample overlap was introducing bias (weak instrument bias or winner’s curse) to our results. We were able to run MR-Lap for the analysis using CRP and GlycA, however the threshold limit set by the package resulted in a loss of all IVs in the IL-6 analysis.

The majority of analyses were unaffected by biases, although the correct IVW effect did differ from the observed IVW effect when investigating the effect of total n-6 PUFAs on GlycA levels and on CRP levels. In both cases, the observed effect estimate was in the same direction as the corrected estimate.

## Potential causal effect of DHA levels on inflammatory markers

### Effect of DHA on outcome CRP levels

There were 45 instrumental variables left after pruning and in the observed analysis there was no effect of DHA on CRP (IVW: 0.05; 95% CI:-0.02, 0.12). The corrected MR-Lap estimate suggested that biases were not affecting our results (p=0.508), and also found no effect of DHA on CRP levels (corrected IVW: 0.05; -0.02, 0.12).

### Effect of DHA on outcome GlycA levels

There were 42 instrumental variables left after pruning and in the observed analysis there was no effect of DHA on GlycA (IVW:-0.04; -0.12, 0.05). The corrected MR-Lap estimate suggested that biases were not affecting the results difference (p=0.06) and also found no effect of DHA on GlycA levels (corrected IVW: -0.03; -0.12, 0.06).

## Potential causal effect of LA levels on inflammatory markers

### Effect of LA on outcome CRP levels

There were 52 instrumental variables left after pruning and in the observed analysis there was no effect of LA on CRP (IVW: 0.01; -0.08, 0.10 ). The corrected MR-Lap estimate suggested that biases were not affected our results (p=0.751) and also found no effect of LA on CRP (corrected IVW: 0.01; -0.09, 0.11).

### Effect of LA on outcome GlycA levels

There were 55 instrumental variables left after pruning and in the observed analysis there was an effect of LA on GlycA (IVW: -0.28; -0.39, -0.18). The corrected MR-Lap estimate suggested that biases were not affected our results (p=0.197) and also found an effect of LA on GlycA levels (corrected IVW: -0.27; -0.39, -0.16).

## Potential causal effect of n-3 levels on inflammatory markers

### Effect of total n-3 PUFAs on outcome CRP levels

There were 45 instrumental variables left after pruning an in the observed analysis there was an effect of n-3 on CRP levels (IVW: 0.08; 0.01, 0.14). The corrected MR-Lap estimate suggested that biases were not affecting our results (p=0.983) and also found an effect of n-3 on CRP (corrected IVW: 0.08; 0.01, 0.14).

### Effect of total n-3 PUFAs on outcome GlycA levels

There were 45 instrumental variables left after pruning and in the observed analysis there was an effect of total n-3 on GlycA levels (IVW: -0.20; -0.32, -0.09). The corrected MR-Lap estimate suggested that biases were not affecting our results (p=0.985) and also found an effect of total n-3 on GlycA (-0.20; -0.32, -0.20).

## Potential causal effect of n-6 levels on inflammatory markers

### Effect of total n-6 PUFAs on outcome CRP levels

There were 62 instrumental variables left after pruning and in the observed analysis there was no effect of n-6 on CRP levels (IVW: 0.05; -0.04, 0.14). The corrected MR-Lap estimate suggested that biases were affecting our results (p=2.52x10-6), but also found no effect of total n-6 on CRP (corrected IVW: 0.05; -0.04, 0.15).

### Effect of total n-6 PUFAs on outcome GlycA levels

There were 61 instrumental variables left after pruning and in the observed analysis there was an effect of n-6 on GlycA levels (IVW: -0.30; -0.40, -0.20). The corrected MR-Lap estimate suggested that biases were affecting our results (p=0.006), but the effect estimate was in the same direction and the corrected estimate also found an effect of total n-6 on GlycA (corrected IVW: -0.29; -0.40, -0.18).

## Results from MR analysis focusing on specific PUFA genes

Estimates using only SNPs identified in the *FADS* gene region indicated that long-chain PUFA concentrations (DHA) increased CRP levels (IVW=0.04; 95% CI=0.01, 0.07), but short chain PUFA concentrations (LA) decreased CRP levels (-0.14; -0.20, -0.07). Estimates when using the SNPs identified to be in the *FADS* gene cluster from the DHA GWAS and the LA GWAS indicated the gene had no effect on GlycA levels (-0.004; -0.03, 0.03 and 0.05; -0.06, 0.13 respectively). Similarly, estimates when using the SNPs identified to be in the *FADS* gene cluster from the DHA GWAS indicated the gene had no effect on IL-6 levels (Wald ratio = -0.01; -0.05, 0.03).

Estimates when using the SNPs identified to be in the *ELOVL2* gene from the DHA GWAS and LA GWAS indicated the gene had no effect on GlycA levels (I-0.11; -0.40, 0.18 and 0.18; -0.08, 0.44 respectively). This was consistent with their effect on CRP levels (-0.16; -0.35,0.03 and 0.02; -0.14, 0.18 respectively) or on IL-6 levels (0.33; -0.62, 1.29 and wald ratio = -0.78; -3.25, 1.69).

## Results from MR-Clust analysis to explore heterogeneity in MR instruments

In the MR-Clust analysis investigating the effect of SNPs identified in the LA GWAS on GlycA levels, all clusters showed evidence of a positive effect. Similarly, when using SNPs identified in the total n-3 PUFA GWAS and SNPs identified in the total n-6 PUFA GWAS, all clusters were positively associated with GlycA levels. The positive effect on all clusters is consistent with the positive effect of the SNPs in the univariable MR analysis. In contrast, when using genetic instruments identified in the DHA GWAS, there were two clusters of five SNPs that were positively associated with GlycA and one cluster of six SNPs that was negatively associated with GlycA (see eTable 26 for a breakdown of the number of SNPs in each cluster).

## MVMR results

When using MVMR to investigate the direct effect of total n-3 PUFAs and total n-6 PUFAs on GlycA there was evidence of heterogeneity across the set of instruments (Q=1504.022, p=<0.001). The overall conditional F-statistic for instrument strength was larger than the conventional threshold of 10 for total n-3 PUFAs (F=67.71) and total n-6 PUFAs (F=99.85) which suggests that the result was not biased by weak instruments.

There was evidence of heterogeneity across the set of instruments used in the MVMR to investigate the direct effect of total n-3 PUFAs and total n-6 PUFAs on CRP (Q=1296.58, p=<0.001). The overall conditional F-statistic for instrument strength in this analysis was larger than the conventional threshold of 10 for total n-3 PUFAs (F=61.20) and total n-6 PUFAs (F=161.65) which suggests that the result was not biased by weak instruments.

When using MVMR to investigate the direct effect of total n-3 PUFAs and total n-6 PUFAs on IL-6 there was no evidence of heterogeneity (Q=10.36, p=0.110). The overall F-statistic for instrument strength was larger than the conventional threshold for 10 for both total n-3 PUFAs (13.55) and total n-6 PUFAs (12.17) which suggests that the result was not biased by weak instruments.

# eReferences

1. Sanderson E, Glymour MM, Holmes MV, et al. Mendelian randomization. *Nature Reviews Methods Primers* 2022; **2**: 1-21.

2. Richmond RC, Davey Smith G. Mendelian Randomization: Concepts and Scope. *Cold Spring Harbor Perspectives in Medicine* 2022; **12**: a040501.

3. Davey Smith G, Ebrahim S. ‘Mendelian randomization’: can genetic epidemiology contribute to understanding environmental determinants of disease? *International journal of epidemiology* 2003; **32**: 1-22.

4. Lawlor DA, Tilling K, Davey Smith G. Triangulation in aetiological epidemiology. *International Journal of Epidemiology* 2017; **45**: 1866-86.

5. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform* 2009; **42**: 377-81.

6. Spratt M, Carpenter J, Sterne JAC, et al. Strategies for Multiple Imputation in Longitudinal Studies. *American journal of epidemiology* 2010; **172**: 478-87.

7. Clayton GL, Borges MC, Lawlor DA. From menarche to menopause: the impact of reproductive factors on the metabolic profile of over 65,000 women. *medRxiv* 2022: <https://doi.org/10.1101/2022.04.17.22273947>

8. Said S, Pazoki R, Karhunen V, et al. Genetic analysis of over half a million people characterises C-reactive protein loci. *Nature Communications* 2022; **13**: 2198.

9. Ahluwalia TS, Prins BP, Abdollahi M, et al. Genome-wide association study of circulating interleukin 6 levels identifies novel loci. *Hum Mol Genet* 2021; **30**: 393-409.

10. Ligthart S, Vaez A, Võsa U, et al. Genome analyses of> 200,000 individuals identify 58 loci for chronic inflammation and highlight pathways that link inflammation and complex disorders. *The American Journal of Human Genetics* 2018; **103**: 691-706.

11. Mounier N. Bias correction for inverse variance weighting Mendelian randomization. *Genetic Epidemiology* 2023; **47**: 314-31.