






Opinion

Vexed causal inferences in nutritional epidemiology—call for genetic help

Pauli Ohukainen ^{1,2}, Jyrki K Virtanen ³, and
Mika Ala-Korpela ^{1,2,4*}

¹Computational Medicine, Faculty of Medicine, University of Oulu and Biocenter Oulu, Oulu, Finland; ²Center for Life Course Health Research, University of Oulu, Oulu, Finland; ³Institute of Public Health and Clinical Nutrition, University of Eastern Finland, Kuopio, Finland and ⁴NMR Metabolomics Laboratory, School of Pharmacy, University of Eastern Finland, Kuopio, Finland

*Corresponding author. Computational Medicine, Center for Life Course Health Research, Faculty of Medicine, University of Oulu, Oulu, Finland. E-mail: mika.ala-korpela@oulu.fi

Editorial decision 23 June 2021; accepted 9 July 2021

Nutritional epidemiology is criticized for its inability to provide plausible information on the causal effects of diet on health and disease outcomes, a key to both scientific understanding and guiding public policies.^{1,2} Although most public dietary guidelines globally have arrived at similar conclusions, they all continue to rely strongly on observational studies.³ It is well known in epidemiology that observational study settings produce the lowest-quality data with perceivable but also unidentifiable confounding and with very limited, if any, opportunities to assess direct causality. However, we would hope for the same reliability for nutritional recommendations as for pharmacological treatments—only substances known to have causal effects should be recommended and used.⁴ Thus, randomized controlled trials (RCTs) might be seen as a favourable strategy to produce reliable causal evidence also in nutritional research.¹ However, RCTs attempting to study the effects of individual foods or nutrients face severe obstacles. For example, a diet always consists of multifactorial and synergistic components, so it is unrealistic for nutritional RCTs to address all the potentially meaningful components of diets, and it may also take years or even decades to have noticeable dietary effects on relevant health outcomes. Low compliance and high dropout rates are also common in nutrition studies if participants are asked to change their typical diets for more than a few months. Blinding, one of

the cornerstones of RCTs, is often impossible if the study requires changes in dietary intakes that cannot be accomplished by supplementation. There is also rarely a placebo group with zero intake of a certain nutrient.

Vitamins provide a classical example of the challenges detailed above.⁵ Multiple observational outcome associations, like for plasma vitamin C (vitC), reflect extensive confounding and, even after adjustment for a range of confounders, residual confounding cannot be ruled out. Measurement errors in the applied confounders and failure to include all relevant confounding factors, e.g. socio-economic and behavioural, from across the life course will be involved and explain the contradicting null results from RCTs.⁵ In recent large-scale Mendelian randomization studies, Dutch researchers looked at the potential causal effects of diet-derived circulating antioxidants (vitamins E and C, retinol, beta-carotene and lycopene) on the risk of coronary heart disease⁶ and stroke.⁷ The results in both studies do not support a protective effect of dietary-derived antioxidant levels on the outcome risk. Thus, it is unlikely that antioxidant supplementation at the population level would be beneficial for the prevention of coronary heart disease or stroke. The total study population in these analyses was extensive, 768 121 individuals with 93 230 cases and 1 065 119 individuals with 77 612 cases for the coronary heart disease⁶ and stroke,⁷ respectively.

The above-mentioned limitations of both observational nutritional epidemiology and related RCTs would call for additional approaches to complement nutritional research to facilitate causal inference. Genetic instrumentation has been suggested as a tool for aiding causal inference in nutritional epidemiology.² In this Opinion, we tackle recent ideas and applications of Mendelian randomization analysis in big data and elucidate the appeal of these new scientific approaches in genetic epidemiology.

Dietary habits and Mendelian randomization analysis

A recent work by Cole *et al.*,⁸ focusing on the genetic background of dietary habits, presents an interesting approach. The authors performed a genome-wide association study (GWAS) on 85 single food intake and related dietary patterns from food frequency questionnaires in the UK Biobank. Over 800 associated genetic loci were identified, allowing the set-up of specific genetic instruments and the application of Mendelian randomization analysis to assess potential causal effects on various health outcomes. The authors reported, for example, that a genetically associated dietary pattern driven by wholemeal vs white bread consumption would be causally influenced by factors correlated with education but would not be causal for coronary artery disease or type 2 diabetes. In fact, instead of demonstrating a causal effect from diet to disease, the findings were interpreted to suggest a reverse causal relationship between coronary artery disease and diet—maybe reflecting a potential behavioural change towards believed healthier food choices due to a heart disease diagnosis. This is analogous to other examples of genetic associations being subject to reverse causality. For example, even though a genetic risk score for coronary heart disease is associated with statin therapy, the true causal pathway is that the disease causes an individual to be prescribed statins, not the other way around.

Mendelian randomization is a form of instrumental variable analysis used to assess causality of exposures using genetic data and it has become increasingly popular in epidemiology over the past decade.^{4,9} A clear distinction should be made between Mendelian randomization and a GWAS, the former being based on the latter. This is also reflected in the interpretation of results: a GWAS provides genetic associations in a general sense but the Mendelian randomization framework assesses potential causal effects.¹⁰ The fundamental principles and prerequisites for univariable Mendelian randomization analysis are exemplified in Figure 1. The three axioms of Mendelian randomization analysis are: (i) the genetic variant(s) must associate with the exposure; but (ii) not with either known or unknown confounders; and (iii) there should be no

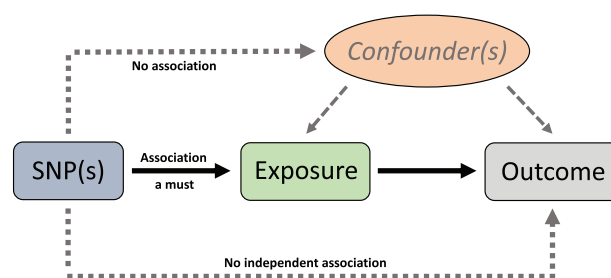


Figure 1 The fundamental principles and prerequisites for univariable Mendelian randomization analysis to estimate causal relationships. There are three key principles in this instrumental variable analysis. The genetic variant (either in isolation or in combination with other variants) must associate with the exposure but must not associate with either known or unknown confounders, and there should be no pathway from the genetic variant(s) to the outcome which does not include the exposure of interest. In two-sample Mendelian randomization analysis, the single nucleotide polymorphism (SNP)-to-exposure estimate is obtained from a dataset separate from that of the SNP-to-outcome estimate. This allows the use of the best existing genome-wide association study (GWAS) for both the exposure (e.g. a circulating biomarker) and the outcome (e.g. coronary heart disease) in rather common situations where a single appropriate dataset is not available. This is a schematic representation and should not be interpreted as a formal directed acyclic graph.

pathway from the genetic variant(s) to the outcome which does not include the exposure of interest. In nutritional epidemiology, however, the interpretation of the genetic component of predominantly environmental traits, such as dietary intake, is complicated and prone to various caveats, as also emphasized by Cole *et al.*⁸ Dietary habits are highly correlated both with each other and with non-dietary traits, suggesting that any single dietary phenotype may represent a broader diet and lifestyle, for example, with confounding links to obesity and socioeconomic status. From a methodological point of view, inevitable genetic pleiotropy and weak instruments complicate the quest for robust findings. It should not be surprising that our preference, e.g. for eating a certain type of bread, can at best be weakly affected by genetics but is heavily dependent on cultural and socioeconomic determinants. Thus, genetic instrumentation and Mendelian randomization analyses might not provide much help in assessing the causality of intake of a single food or dietary pattern on outcomes in a univariable setting.⁸ However, there can be specific cases in which a genetic variant provides a good instrument for mimicking the dietary intake of a single nutrient. The observational¹¹ and RCT evidence¹² in relation to vitC has been complemented with a Mendelian randomization study.¹³ Variation in the solute carrier family 23 member 1 (*SLC23A1*) gene is robustly associated with circulating vitC levels,¹⁴ making it a rare but excellent instrumental variable for studying potential long-term causal effects of this nutrient.

The complex relationships between genetics, dietary exposures and potential causal pathways with health outcomes were recently elaborated by Pirastu *et al.*¹⁵ They developed a statistical genetics framework, involving multivariable Mendelian randomization analysis, to better account for confounding and collider bias in order to increase the robustness of causal analyses.¹⁶ They demonstrated that genetic associations with dietary traits are likely also affected by reverse causality, thus giving a potential explanation of the heterogeneity of genetic correlations and genome-wide associations in different populations. The authors noted that when considering the effects of foods on health, the genetic evidence would support the importance of dietary patterns rather than single foods or nutrients. This method has also recently been applied to study the potential causal associations of dietary patterns on over a hundred circulating biomarkers analysed by nuclear magnetic resonance (NMR) metabolomics.¹⁷ These Mendelian randomization analyses identified more than 400 potentially causal links between food and biomarkers with replication of some previous findings, e.g. increased oily fish consumption and higher circulating docosahexaenoic acid (DHA) concentrations. Among new causal findings were various food effects on apolipoprotein B-containing lipoprotein particles.¹⁷

In another recent study using the UK Biobank data, Dashti *et al.* analysed the potential causal effects of morning-evening diurnal preferences on food intake.¹⁸ Their results suggest that a morning diurnal preference causes increased intake of higher-quality foods and decreased intake of lower-quality foods. These results may reflect temporality in the consumption of foods and thus warrant the assessment of diurnal preferences in public health and in relation to dietary advice.

Genetic pleiotropy

It is important to note that genetic pleiotropy is an established and pervasive characteristic of the human genome. Consequently, we should not think as simply as ‘one gene, one function, one trait’, but be aware that many single genetic variants influence a plethora of different traits.¹⁹ A violation of the prerequisites for Mendelian randomization analyses is the so-called horizontal pleiotropy that refers to genetic variants influencing two traits via independent pathways. However, a genetic variant influencing other traits on the same pathway (so-called vertical pleiotropy) is not a direct violation of the axioms of Mendelian randomization analysis. The obstacle of (horizontal) pleiotropy has been tackled extensively in the literature and readers are referred to a recent review by Hemani *et al.*²⁰ for further details.

The studies by Papadimitriou *et al.*²¹ and Burrows *et al.*²² are recent exemplars of how to control for genetic pleiotropy via extensive sensitivity analyses in nutritional Mendelian randomization applications. Papadimitriou *et al.* tackled the inconsistencies in observational associations between the consumption or circulating concentrations of micronutrients and breast cancer risk. Of the 11 micronutrients studied, higher concentrations of magnesium were causally related to the increased risk of breast cancer.²¹ Burrows *et al.* challenged the contradictory observational evidence on associations between folate and the risk of several common cancers. Their results gave little evidence that serum folate would be causal for pan-cancer or various site-specific cancers, suggesting that increasing levels of circulating folate is unlikely to lead to population-wide increase in cancer risk.²²

Mediating biomarkers and multivariable causal mediation analysis

In spite of the above-mentioned methodological complexities, and even some unfounded resistance expressed by some prominent researchers,²³ we anticipate that new developments with multivariable Mendelian randomization (MVMR) analysis are of particular value in nutritional epidemiology. In MVMR, many scenarios are possible between exposures, mediators and an outcome.²⁴ This approach also retains the benefits of using genetic instruments for causal inference, thus avoiding bias due to confounding while allowing for estimation of the different effects required for mediation analysis. As we demonstrate in Figure 2, in the assessment of the effects of various lifestyle-related dietary factors, we can consider several likely physiological and molecular mediators between the primary exposures and the outcome.²⁴ Both direct and indirect causal associations are likely in many nutritional situations. Here we would like to focus on the potential of Mendelian randomization analyses to support nutritional epidemiology, particularly with respect to the robustness of the interpretations of potential causal consequences of dietary observations. Reliable Mendelian randomization analyses are already available for many biomarkers to aid causal interpretations in nutritional studies.^{4,9,23}

Table 1 provides an exemplary list of markers affected by diet, related Mendelian randomization evidence and RCT data. For example, systolic blood pressure is known to be affected by diet⁵⁶ and its causality for cardiovascular disease (CVD) is supported by both Mendelian randomization analyses and RCTs,^{54,55} endorsing it as a valid mediating biomarker in nutritional epidemiology. Conversely, although low selenium status was observationally associated with higher risk of prostate cancer,³⁹ subsequent

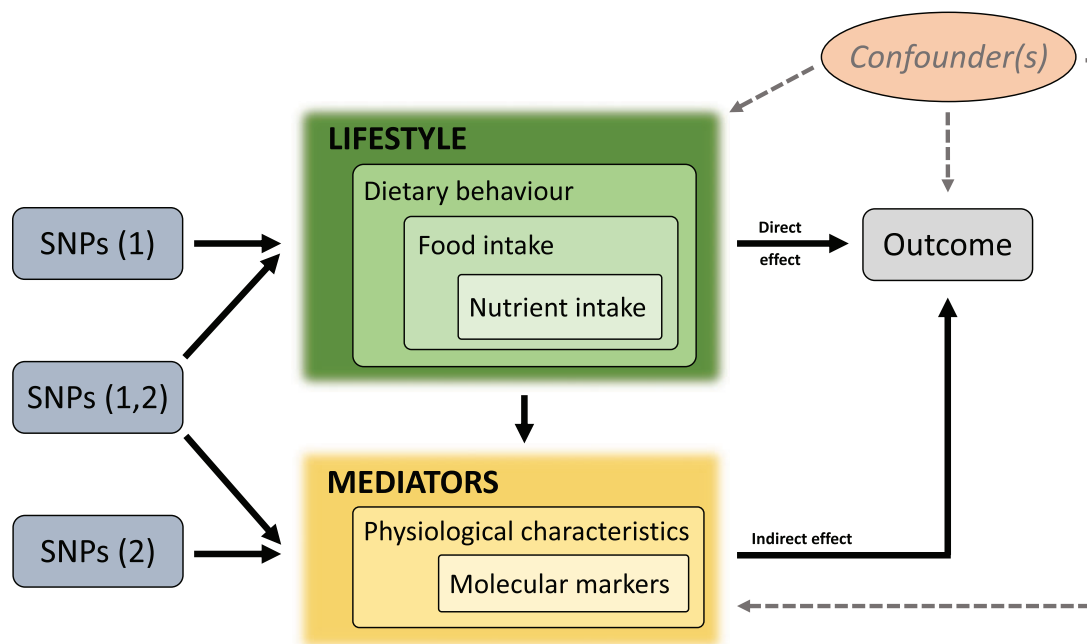


Figure 2 A general schematic illustration of multivariable Mendelian randomization analysis in a mediation scenario to assess direct causal effects of lifestyle-related dietary factors and their indirect causal effects as mediated by physiological and molecular exposures. Single nucleotide polymorphisms (SNPs) (1) refers to the genetic variants affecting the lifestyle-related dietary factors; SNPs (2) refers to the genetic variants affecting the physiological and molecular exposures; and SNPs (1,2) refers to the genetic variants affecting both exposures. Consequently, the causal effects estimated by univariable and multivariable Mendelian randomization analysis can differ. Univariable estimates represent the total causal effect of the exposure on the outcome, whereas multivariable estimates constitute the direct causal effect of each exposure on the outcome. This is a schematic representation of a common situation in nutritional epidemiology research and should not be interpreted as a formal directed acyclic graph.

RCTs with selenium supplementation were discontinued due to increased risk,⁴¹ and the same outcome was later reported in a Mendelian randomization analysis.⁴⁰ In general, it would be essential to realize that observational epidemiology cannot elucidate causality but, when Mendelian randomization analyses have been possible, they have usually produced consistent results with RCTs.²³

A particularly fruitful area of research may lie at an intersection of RCTs and Mendelian randomization analysis. Instead of attempting to stretch nutritional trial durations long enough to cover clinical endpoints, a more practical approach may be to first conduct a smaller and shorter RCT followed by Mendelian randomization. This ‘two-step randomization’ approach has been performed to investigate potential clinical benefits of lycopene supplementation on prostate cancer risk.⁵⁷ First, a feasibility trial of lycopene supplementation⁵⁸ was done to determine the systemic metabolic effects of the intervention. Circulating pyruvate concentrations were lowered and the long-term causal effects of this were then investigated in a Mendelian randomization framework, suggesting a lowered prostate cancer risk.⁵⁷ The logic in this clinical interpretation relies on the assumption that the supplementation effects observed in the short-term intervention step would be sustained long

term. On the other hand, the results from the Mendelian randomization with respect to the causal role of circulating pyruvate for prostate cancer are valid regardless, as they indicate the lifelong genetic effects of the metabolite in question.^{4,9}

Perils of hasty causal expectations—‘good cholesterol’ and vitamin D

Recent years have provided a cautionary exemplar of overreliance on observational epidemiology, namely circulating high-density lipoprotein cholesterol (HDL-C), the so-called ‘good cholesterol’ for many years. Since the early days of cardiovascular research, low HDL-C has been seen as an unquestionable risk factor for CVD, but the situation has become intricate with modern genetics.⁴ Coherent observational findings led to a hypothesis that any intervention elevating circulating HDL-C would lower the risk of CVD. However, multiple clinical trials testing this hypothesis have either failed or demonstrated benefit unrelated to raising circulating HDL-C concentrations.^{59,60} Consistent with these null trial outcomes, various Mendelian randomization analyses have also indicated that HDL-C would not be causal for CVD.⁶¹ The latest evidence is also questioning the causality of apolipoprotein A-I (apoA-I), together

Table 1 Selected examples of associations between diet-related biomarkers and various disease outcomes via observational study settings, Mendelian randomization (MR) analysis and randomized controlled trials (RCTs)

Diet or supplementation biomarker	Disease outcome	Observational evidence	Mendelian randomization evidence for causality	RCT evidence for causality	Interpretation
Vitamin D	Cardiovascular disease	Inverse association with cardiovascular mortality ²⁵	No causal relationship for coronary artery disease ²⁶	No benefit in a meta-analysis of 21 clinical trials ²⁷	MR analyses and RCTs are consistent: elevating serum vitamin D by supplementation unlikely to reduce cardiovascular mortality
	Colorectal cancer	Inverse association ²⁸	No causal relationship ²⁹	No reduction of risk by supplementation ^{30,31}	MR analyses and RCTs are consistent: elevating serum vitamin D by supplementation unlikely to reduce risk
	Breast cancer	Inverse ³² or no association ³³	No causal relationship ²⁹	No reduction of risk by supplementation ³⁴	MR analyses and RCTs are consistent: vitamin D unlikely associated with risk; RCTs are contraindicated
	Prostate cancer	No association ³³	No causal relationship ²⁹	No reduction of risk by supplementation ³¹	All studies are consistent: vitamin D unlikely associated with risk; RCTs are contraindicated
Homocysteine	Stroke	Positive association ³⁵	A causal relationship ³⁶	Meta-analysis of 25 trials of homocysteine lowering (with vitamin B) ³⁷ indicates a causal effect.	All information is consistent: lowering serum homocysteine likely to reduce risk of stroke
	Ischaemic heart disease	Positive association ³⁵	No causal relationship ³⁸	Meta-analysis of 10 placebo-controlled studies shows no benefit ³⁸	MR studies and RCTs are consistent: lowering serum homocysteine unlikely to reduce risk of ischaemic heart disease
Selenium	Prostate cancer	Toenail selenium inversely associated with risk ³⁹	A causal relationship for blood selenium ⁴⁰	Trial of selenium (and vitamin E) supplementation discontinued due to increased risk of prostate cancer ⁴¹	MR analyses and RCTs are consistent: increasing selenium intake likely to increase risk
HbA1c	Cardiovascular disease	Positive association with higher risk of mortality and myocardial infarction in patients with coronary artery disease ⁴²	A causal relationship ⁴³	SGLT2i reduces HbA1c and CVD mortality in type 2 diabetic patients ⁴⁴	All information is consistent: lowering HbA1c likely to prevent CVD
Vitamin E	Coronary heart disease	Inverse association ⁴⁵	A causal relationship ⁴⁶	No benefit in high-risk individuals ¹⁰	MR analyses and RCTs are consistent: No protective effect of supplementation and potential safety concerns
IL-6	Rheumatoid arthritis	Positive association ⁴⁷	An inverse causal relationship ⁴⁸	Inhibition of IL-6 receptor with tocilizumab reduces disease activity and improves function ⁴⁹	MR analyses and RCTs are consistent: inhibition of IL-6 pathway is beneficial

(Continued)

Table 1 Continued

Diet or supplementation biomarker	Disease outcome	Observational evidence	Mendelian randomization evidence for causality	RCT evidence for causality	Interpretation
Calcium	Cardiovascular disease	Positive association ⁵⁰	A causal relationship ⁵¹	Supplementation increases risk of vascular outcomes ⁵²	All information is consistent: serum calcium positively associated with risk of CVD
Blood pressure	Cardiovascular disease	Positive association ⁵³	A causal relationship ⁵⁴	Lowering blood pressure prevents cardiovascular disease ⁵⁵	All information is consistent: lowering blood pressure is beneficial for prevention and treatment of CVD
Vitamin C	Cardiovascular disease	Inverse association ¹¹	No causal relationship ¹³	Supplementation does not reduce risk of cardiovascular disease ¹²	MR analyses and RCTs are consistent: supplementing with vitamin C does not reduce CVD

Studies listed do not represent a chronological accumulation or the totality of a specific type of evidence. In most cases, RCT evidence precedes genetic evidence. There are also often multiple observational studies for the same exposure and outcome, but only one representative work is cited. A narrative reference tracking search strategy was used to identify examples where all three types of studies addressing the same exposure were available.

CVD, cardiovascular disease; HbA1c, glycated haemoglobin; IL-6, interleukin-6; SGLT2i, sodium-glucose cotransporter 2 inhibitor.

with some of the functional features of HDL particles—a bad omen for the ongoing apoA-I infusion trials.⁶²

In nutritional epidemiology, a story with similar features is unfolding around vitamin D (vitD). The observational evidence for the association between low circulating vitD and multiple disease outcomes is strong and long-lived. However, recent RCTs of vitD supplementation in the prevention of cancer³⁰ and cardiovascular disease³¹ have led to null results (Table 1). Almost simultaneously, Mendelian randomization analyses have suggested that vitD would not be causal for schizophrenia, prostate cancer or bone mineral density but may be for ovarian cancer, multiple sclerosis and Alzheimer’s disease.⁶³ RCTs of vitD supplementation are complicated by issues around the optimal dosing regimen, study duration, baseline vitD status and non-supplement sources of vitD. Ethical considerations are also paramount. For example, how long can we monitor vitD-deficient participants? Due to numerous difficulties in conducting an optimal nutritional trial, fundamental questions about causality can be excessively difficult to answer. Serum vitD is an example of an exposure that can have multiple upstream drivers, including diet. Nevertheless, despite the inherent difficulties in directly assessing the potential causal effects of dietary vitD, we can use serum vitD as an exposure in Mendelian randomization analyses. Even though population determinants of vitamin D status are numerous, circulating vitD is modified by supplementation and the well-characterized genetic regulation allows the build-up of an appropriate genetic instrument.⁶⁴

Bidirectional Mendelian randomization for vitamin D and multiple diseases

A recent genome-wide association study of vitD concentration in the UK Biobank (which identified 143 independent loci), together with bidirectional Mendelian randomization analyses, is an elegant exemplar of how the causal role of a biomarker can be extensively evaluated.⁶⁵ After paying particular attention to accounting for genetic pleiotropy, an overall conclusion of these analyses was that there was no robust evidence that vitD would have causal effects on body mass index and multiple disease phenotypes. In addition, Revez *et al.*⁶⁵ concluded that observational epidemiological links between vitD and psychiatric disorders mostly reflect confounding and/or reverse causation. In general, this recent work demonstrates the possibilities of Mendelian randomization and also suggests that many phenotypes do appear to have causal effects on vitD concentration, thereby emphasizing an important role of reverse causation for this particular biomarker and

reminding us to be sceptical in interpreting evidence arising from observational epidemiological studies.

Dietary fats and cardiovascular disease

The long-term nutritional research on the potential direct health effects of dietary fats is an exemplar of the vast difficulties in acquiring robust causal evidence. The early research suggested that the reduction of saturated fats, especially when replaced with polyunsaturated fats, would be beneficial for cardiometabolic health.⁶⁶ Serum total cholesterol was suggested as a potential mediator of this effect, leading to the classic ‘diet-heart hypothesis’. Total cholesterol was eventually replaced by low-density lipoprotein cholesterol (LDL-C) but the hypothesis remained essentially the same. Controlled feeding studies in metabolic wards have established that consuming saturated fats leads to increased LDL-C concentrations,⁶⁷ a finding also supported by mechanistic studies.⁶⁸ Nevertheless, rigorous testing of this hypothesis has proved challenging and the present-day dietary guidelines have been questioned, for example, due to the lack of independent association between dietary saturated fat and CVD in more recent epidemiological studies⁶⁹ and the lack of benefit when replacing saturated fats with omega-6 fatty acids in RCTs.⁷⁰ Fundamental issues of confounding from the food matrix of individual foods as well as the totality of diet and varying effects of different species of saturated fatty acids also apply.

Recent observational findings from the large-scale Prospective Urban Rural Epidemiology study, with over 125 000 participants from 18 countries from five continents, have been interpreted to be at odds with current recommendations to reduce total fat and saturated fats.^{71,72} The results reveal that nutrients have varying effects on different circulating lipoprotein risk factors. The authors argue that predicting the net clinical effect based on considering only the effects of nutrient intake on LDL-C would not be reliable in projecting the effects of diet on CVD events or on total mortality. The situation is complicated by the fact that some lipid biomarkers are causal for cardiometabolic outcomes (LDL-C, apolipoprotein B and triglycerides)⁷³ but, as discussed above, the recent genetic evidence strongly suggests, in contradiction to the observational associations, that HDL-C and apoA-I would not be causal.^{61,62} Within the caveats discussed above in relation to epidemiological studies as well as nutritional RCTs, it also needs to be kept in mind that whatever the effects of fats on plasma lipids, we would need to remain cautious in causal interpretations between fats and disease outcomes. Nevertheless, the most recent meta-analysis of RCT evidence does suggest that reducing saturated fat intake for at

least 2 years causes a potentially important reduction in combined cardiovascular events.⁷⁴ The controversy is therefore likely to remain and, alas, Mendelian randomization analysis cannot help in this case since there is no specific biomarker for saturated fat intake (serum saturated fatty acids are not a reliable proxy of dietary saturated fatty acids).

A partial solution to the above-mentioned instrument problem has been recently provided by a GWAS of relative macronutrient intake.⁷⁵ Implementing these single nucleotide polymorphisms (SNPs) in a two-sample Mendelian randomization setting, Park *et al.* were able to assess the causal associations of relative macronutrient intake on the risk of chronic kidney disease (CKD).⁷⁶ Their results suggested that, for a given level of total calorie intake, the composition of macronutrient intake would causally affect the risk of CKD. Thus, reducing relative fat intake and increasing relative protein intake may causally reduce the risk of CKD in the general population.

Mediterranean diet and cardiovascular disease

The Mediterranean diet is generally seen favourable for cardiometabolic health. In a recent study, Li and co-workers developed a quantitative score for the adherence to a traditional Mediterranean diet and further linked the score to a systemic metabolic signature.⁷⁷ Although the association between the Mediterranean diet score and the metabolic signature was weak, this approach allowed the application of Mendelian randomization analysis via the genetic instrumentation of the metabolic signature. The genetic component of the signature was inversely associated with risk of coronary heart disease and stroke. Several risk factors, including blood lipids, systolic blood pressure and diabetes, showed weak mediating effects in these Mendelian randomization analyses. In general, the quantification of habitual diet (with or without related biomarker signatures) may be a more realistic concept to assess causal pathways from diet to disease than focusing on individual foods and nutrients. Further genetic support for the favourable effects of the Mediterranean diet, high in vegetables, comes from a recent study by Park *et al.*, suggesting that higher vegetable intake may be causally associated with better kidney function.⁷⁸

Better appreciation of causality—better quality nutritional epidemiology

Determining reliable and specific biomarkers for food intake is extremely challenging, if not impossible. It should also be noted that diet naturally has effects on multiple

biomarkers and thus a single one would be highly unlikely to capture the physiological effects successfully. However, with Mendelian randomization analysis we now have a potentially powerful tool to assess causality of various biomarkers. Dietary behaviours and individual nutrients are most likely linked to several biological mediators in the human (patho)physiology. Although we may never know or be able to measure all of them, striving to identify causal mediators allows us to make stronger causal inferences for dietary exposures. If none are available or instrumental variable analyses disagree with the observational findings, wider exploration of alternative hypotheses, including methodological shortcomings related to study designs, are warranted. Overall, we highlight the importance of appraising known causal biomarkers before any causal interpretations of observational dietary data.

Food and nutrient intakes are largely influenced by environmental factors with minimal or no genetic contribution. However, when it comes to dietary behaviours, the role of genetics is plausible.⁸ Due to the inherent complexity related to the interplay between diet, genes and the environment, it is challenging to perform robust univariable Mendelian randomization analyses (Figure 1) to assess potential causality of primary dietary factors, including dietary behaviours. However, turning towards mediation analyses in the new framework of multivariable Mendelian randomization (Figure 2) appears to offer a fair amount of promise for assessing unknown potentially causal relations, mediation and pathways. In addition to strengthening overall aetiological understanding, this approach could be highly valuable in determining which RCTs would be the most likely to yield useful information.

Funding

PO is supported by the Emil Aaltonen Foundation. MAK is supported by a research grant from the Sigrid Juselius Foundation, Finland.

Acknowledgement

We sincerely thank George Davey Smith (University of Bristol, UK) for critical notions and incomparable substance advice.

Conflict of interest

None declared.

References

- Ioannidis JPA. The challenge of reforming nutritional epidemiologic research. *JAMA* 2018;320:969–70.
- Davey Smith G. Use of genetic markers and gene-diet interactions for interrogating population-level causal influences of diet on health. *Genes Nutr* 2011;6:27–43.
- Herforth A, Arimond M, Álvarez-Sánchez C, Coates J, Christianson K, Muehlhoff E. A global review of food-based dietary guidelines. *Adv Nutr* 2019;10:590–605.
- Holmes MV, Ala-Korpela M, Davey Smith G. Mendelian randomization in cardiometabolic disease: challenges in evaluating causality. *Nat Rev Cardiol* 2017;14:577–90.
- Lawlor DA, Davey Smith G, Bruckdorfer KR, Kundu D, Ebrahim S. Those confounded vitamins: what can we learn from the differences between observational versus randomised trial evidence? *Lancet* 2004;363:1724–27.
- Luo J, Le Cessie S, van Heemst D, Noordam R. Diet-derived circulating antioxidants and risk of coronary heart disease. *J Am Coll Cardiol* 2021;77:45–54.
- Martens LG, Luo J, van Dijk KW, Jukema JW, Noordam R, van Heemst D. Diet-derived antioxidants do not decrease the risk of ischemic stroke: a Mendelian Randomization Study in over 1 million participants. *medRxiv* 2021;2021.03.19.21253362.
- Cole JB, Florez JC, Hirschhorn JN. Comprehensive genomic analysis of dietary habits in UK Biobank identifies hundreds of genetic associations. *Nat Commun* 2020;11:1467.
- Davey Smith G, Hemani G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Hum Mol Genet* 2014;23:R89–98.
- Kraft P, Chen H, Lindström S. The use of genetic correlation and Mendelian randomization studies to increase our understanding of relationships between complex traits. *Curr Epidemiol Rep* 2020;7:104–12.
- Khaw K-T, Bingham S, Welch A *et al.* Relation between plasma ascorbic acid and mortality in men and women in EPIC-Norfolk prospective study: a prospective population study. *Lancet* 2001;357:657–63.
- Collins R, Armitage J, Parish S, Sleight P, Peto R. MRC/BHF Heart Protection Study of antioxidant vitamin supplementation in 20 536 high-risk individuals: a randomised placebo-controlled trial. *Lancet* 2002;360:23–33.
- Wade KH, Forouhi NG, Cook DG *et al.* Variation in the SLC23A1 gene does not influence cardiometabolic outcomes to the extent expected given its association with l-ascorbic acid. *Am J Clin Nutr* 2015;101:202–209.
- Timpson NJ, Forouhi NG, Brion M-J *et al.* Genetic variation at the SLC23A1 locus is associated with circulating concentrations of l-ascorbic acid (vitamin C): evidence from 5 independent studies with >15,000 participants. *Am J Clin Nutr* 2010;92:375–82.
- Pirastu N, McDonnell C, Grzeszkowiak EJ *et al.* Using genetic variation to disentangle the complex relationship between food intake and health outcomes. *bioRxiv* 2020;829952.
- Day FR, Loh P-R, Scott RA, Ong KK, Perry JRB. A robust example of collider bias in a genetic association study. *Am J Hum Genet* 2016;98:392–93.
- Taba N, Valge H-K, Metspalu A *et al.* Mendelian randomization identifies the potential causal impact of dietary patterns on circulating blood metabolites. *bioRxiv* 2020;2020.10.09.332924.
- Dashti HS, Chen A, Daghlas I, Saxena R. Morning diurnal preference and food intake: a Mendelian randomization study. *Am J Clin Nutr* 2020;112:1348–57.
- Visscher PM, Wray NR, Zhang Q *et al.* 10 years of GWAS discovery: biology, function, and translation. *Am J Hum Genet* 2017;101:5–22.

20. Hemani G, Bowden J, Davey Smith G. Evaluating the potential role of pleiotropy in Mendelian randomization studies. *Hum Mol Genet* 2018;**27**:R195–208.
21. Papadimitriou N, Dimou N, Gill D *et al.* Genetically predicted circulating concentrations of micronutrients and risk of breast cancer: a Mendelian randomization study. *Int J Cancer* 2021;**148**:646–53.
22. Burrows K, Kazmi N, Haycock P, Tsilidis KK, Martin RM, Lewis SJ. Mendelian randomisation study exploring the associations of serum folate with pan and site-specific cancers. *bioRxiv* 2020;762138.
23. Davey Smith G, Holmes MV, Davies NM, Ebrahim S. Mendel's laws, Mendelian randomization and causal inference in observational data: substantive and nomenclatural issues. *Eur J Epidemiol* 2020;**35**:99–111.
24. 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**:713–27.
25. Dobnig H, Pilz S, Scharnagl H *et al.* Independent association of low serum 25-hydroxyvitamin D and 1,25-dihydroxyvitamin D levels with all-cause and cardiovascular mortality. *Arch Intern Med* 2008;**168**:1340–49.
26. Manousaki D, Mokry LE, Ross S, Goltzman D, Richards JB. Mendelian randomization studies do not support a role for vitamin D in coronary artery disease. *Circ Cardiovasc Genet* 2016;**9**:349–56.
27. Barbarawi M, Kheiri B, Zayed Y *et al.* Vitamin D supplementation and cardiovascular disease risks in more than 83 000 individuals in 21 randomized clinical trials. *JAMA Cardiol* 2019;**4**:765–76.
28. McCullough ML, Zoltick ES, Weinstein SJ *et al.* Circulating vitamin D and colorectal cancer risk: an international pooling project of 17 cohorts. *J Natl Cancer Inst* 2019;**111**:158–69.
29. Dimitrakopoulou VI, Tsilidis KK, Haycock PC *et al.*; GAMEON Network (CORRECT, DRIVE, ELLIPSE, FOCI-OCAC, TRICL-ILCCO). Circulating vitamin D concentration and risk of seven cancers: Mendelian randomisation study. *BMJ* 2017;**359**:j4761.
30. Scragg R, Khaw K-T, Toop L *et al.* Monthly high-dose vitamin D supplementation and cancer risk: A post hoc analysis of the vitamin D assessment randomized clinical trial. *JAMA Oncol* 2018;**4**:e182178.
31. Manson JE, Cook NR, Lee I-M *et al.*; VITAL Research Group. Vitamin D supplements and prevention of cancer and cardiovascular disease. *N Engl J Med* 2019;**380**:33–44.
32. Chen P, Hu P, Xie D, Qin Y, Wang F, Wang H. Meta-analysis of vitamin D, calcium and the prevention of breast cancer. *Breast Cancer Res Treat* 2010;**121**:469–77.
33. Gandini S, Boniol M, Haukka J *et al.* Meta-analysis of observational studies of serum 25-hydroxyvitamin D levels and colorectal, breast and prostate cancer and colorectal adenoma. *Int J Cancer* 2011;**128**:1414–24.
34. Zhou L, Chen B, Sheng L, Turner A. The effect of vitamin D supplementation on the risk of breast cancer: a trial sequential meta-analysis. *Breast Cancer Res Treat* 2020;**182**:1–8.
35. Homocysteine Studies Collaboration. Homocysteine and risk of ischemic heart disease and stroke. *JAMA* 2002;**288**:2015–22.
36. Casas J, Bautista L, Smeeth L, Sharma P, Hingorani A. Homocysteine and stroke: evidence on a causal link from Mendelian randomisation. *Lancet* 2005;**365**:224–32.
37. Marti-Carvajal AJ, Solà I, Lathyris D, Dayer M. Homocysteine-lowering interventions for preventing cardiovascular events. *Cochrane Database Syst Rev* 2017;**8**:CD006612.
38. Clarke R, Bennett DA, Parish S *et al.*; for the MTHFR Studies Collaborative Group. Homocysteine and Coronary Heart Disease: Meta-analysis of MTHFR Case-Control Studies, Avoiding Publication Bias. *PLoS Med* 2012;**9**:e1001177.
39. Yoshizawa K, Willett WC, Morris SJ *et al.* Study of Prediagnostic Selenium Level in Toenails and the Risk of Advanced Prostate Cancer. *J Natl Cancer Inst* 1998;**90**:1219–24.
40. Yarmolinsky J, Bonilla C, Haycock PC *et al.*; PRACTICAL Consortium. Circulating selenium and prostate cancer risk: a Mendelian randomization analysis. *J Natl Cancer Inst* 2018;**110**:1035–38.
41. Klein EA, Thompson IM, Tangen CM *et al.* Vitamin E and the risk of prostate cancer: updated results of the selenium and vitamin E cancer prevention trial (SELECT). *JAMA* 2011;**306**:1549–56.
42. Geng J, Zhang Y, Wang B, Xie J, Xu B, Li J. Glycosylated hemoglobin levels and clinical outcomes in nondiabetic patients with coronary artery disease. *Medicine (Baltimore)* 2017;**96**:e6784.
43. Leong A, Chen J, Wheeler E *et al.* Mendelian randomization analysis of hemoglobin A 1c as a risk factor for coronary artery disease. *Diabetes Care* 2019;**42**:1202–1208.
44. Zinman B, Wanner C, Lachin JM *et al.*; EMPA-REG OUTCOME Investigators. Empagliflozin, cardiovascular outcomes, and mortality in type 2 diabetes. *N Engl J Med* 2015;**373**:2117–28.
45. Rimm EB, Stampfer MJ, Ascherio A, Giovannucci E, Colditz GA, Willett WC. Vitamin E consumption and the risk of coronary heart disease in men. *N Engl J Med* 1993;**328**:1450–56.
46. Wang T, Xu L. Circulating vitamin E levels and risk of coronary artery disease and myocardial infarction: a Mendelian randomization study. *Nutrients* 2019;**11**:2153.
47. Madhok R, Crilly A, Watson J, Capell HA. Serum interleukin 6 levels in rheumatoid arthritis: correlations with clinical and laboratory indices of disease activity. *Ann Rheum Dis* 1993;**52**:232–34.
48. Li B, Xiao Y, Xing D, Ma X, Liu J. Circulating interleukin-6 and rheumatoid arthritis. *Medicine* 2016;**95**:e3855.
49. Singh JA, Beg S, Lopez-Olivio MA. Tocilizumab for rheumatoid arthritis: a Cochrane systematic review. *J Rheumatol* 2011;**38**:10–20.
50. Reid IR, Gamble GD, Bolland MJ. Circulating calcium concentrations, vascular disease and mortality: a systematic review. *J Intern Med* 2016;**279**:524–40.
51. Larsson SC, Burgess S, Michaëlsson K. Association of genetic variants related to serum calcium levels with coronary artery disease and myocardial infarction. *JAMA* 2017;**318**:371–80.
52. Bolland MJ, Barber PA, Doughty RN *et al.* Vascular events in healthy older women receiving calcium supplementation: randomised controlled trial. *BMJ* 2008;**336**:262–66.
53. Lewington S, Clarke R, Qizilbash N, Peto R, Collins R; Prospective Studies Collaboration. Age-specific relevance of

- usual blood pressure to vascular mortality: a meta-analysis of individual data for one million adults in 61 prospective studies. *Lancet* 2002;360:1903–13.
54. Ference BA, Julius S, Mahajan N, Levy PD, Williams KA, Flack JM. Clinical effect of naturally random allocation to lower systolic blood pressure beginning before the development of hypertension. *Hypertension* 2014;63:1182–88.
 55. Brunström M, Carlberg B. Association of blood pressure lowering with mortality and cardiovascular disease across blood pressure levels: a systematic review and meta-analysis. *JAMA Intern Med* 2018;178:28–36.
 56. Bazzano LA, Green T, Harrison TN, Reynolds K. Dietary approaches to prevent hypertension. *Curr Hypertens Rep* 2013;15:694–702.
 57. Sandhu MR, Beynon R, Richmond R *et al.* Two-step randomisation: applying the results of small feasibility studies of interventions to large-scale Mendelian randomisation studies to robustly infer causal effects on clinical endpoints. *Preprints* 2019;2019100276.
 58. Beynon RA, Richmond RC, Santos Ferreira DL *et al.*; PRACTICAL consortium. Investigating the effects of lycopene and green tea on the metabolome of men at risk of prostate cancer: the ProDiet randomised controlled trial. *Int J Cancer* 2019;144:1918–28.
 59. Bowman L, Hopewell JC, Chen F *et al.*; The HPS3/TIMI55–REVEAL Collaborative Group. Effects of anacetrapib in patients with atherosclerotic vascular disease. *N Engl J Med* 2017;377:1217–27.
 60. Kettunen J, Holmes MV, Allara E *et al.* Lipoprotein signatures of cholesteryl ester transfer protein and HMG-CoA reductase inhibition. *PLoS Biol* 2019;17:e3000572.
 61. Richardson TG, Sanderson E, Palmer TM *et al.* Evaluating the relationship between circulating lipoprotein lipids and apolipoproteins with risk of coronary heart disease: a multivariable Mendelian randomisation analysis. *PLoS Med* 2020;17:e1003062.
 62. Karjalainen MK, Holmes MV, Wang Q *et al.* Apolipoprotein A-I concentrations and risk of coronary artery disease: a Mendelian randomization study. *Atherosclerosis* 2020;299:56–63.
 63. Bouillon R, Marcocci C, Carmeliet G *et al.* Skeletal and extraskeletal actions of vitamin D: current evidence and outstanding questions. *Endocr Rev* 2019;40:1109–51.
 64. Berry DJ, Vimalaswaran KS, Whittaker JC, Hingorani AD, Hyppönen E. Evaluation of genetic markers as instruments for Mendelian randomization studies on vitamin D. *PLoS One* 2012;7:e37465.
 65. Revez JA, Lin T, Qiao Z *et al.* Genome-wide association study identifies 143 loci associated with 25 hydroxyvitamin D concentration. *Nat Commun* 2020;11:1647.
 66. Virtanen J. Randomized trials of replacing saturated fatty acids with n-6 polyunsaturated fatty acids in coronary heart disease prevention: not the gold standard? *Prostaglandins Leukot Essent Fatty Acids* 2018;133:8–15.
 67. Clarke R, Frost C, Collins R, Appleby P, Peto R. Dietary lipids and blood cholesterol: quantitative meta-analysis of metabolic ward studies. *BMJ* 1997;314:112.
 68. Woollett LA, Spady DK, Dietschy JM. Saturated and unsaturated fatty acids independently regulate low density lipoprotein receptor activity and production rate. *J Lipid Res* 1992;33:77–88.
 69. Siri-Tarino PW, Sun Q, Hu FB, Krauss RM. Meta-analysis of prospective cohort studies evaluating the association of saturated fat with cardiovascular disease. *Am J Clin Nutr* 2010;91:535–46.
 70. Hamley S. The effect of replacing saturated fat with mostly n-6 polyunsaturated fat on coronary heart disease: a meta-analysis of randomised controlled trials. *Nutr J* 2017;16:30.
 71. Mente A, Dehghan M, Rangarajan S *et al.* Association of dietary nutrients with blood lipids and blood pressure in 18 countries: a cross-sectional analysis from the PURE study. *Lancet Diabetes Endocrinol* 2017;5:774–87.
 72. Dehghan M, Mente A, Zhang X *et al.* Associations of fats and carbohydrate intake with cardiovascular disease and mortality in 18 countries from five continents (PURE): a prospective cohort study. *Lancet* 2017;390:2050–62.
 73. Ala-Korpela M. The culprit is the carrier, not the loads: cholesterol, triglycerides and apolipoprotein B in atherosclerosis and coronary heart disease. *Int J Epidemiol* 2019;48:1389–92.
 74. Hooper L, Martin N, Jimoh OF, Kirk C, Foster E, Abdelhamid AS. Reduction in saturated fat intake for cardiovascular disease. *Cochrane Database Syst Rev* 2020;5:CD011737.
 75. Meddens SFW, de Vlaming R, Bowers P *et al.* Genomic analysis of diet composition finds novel loci and associations with health and lifestyle. *Mol Psychiatry* 2020; doi:10.1038/s41380-020-0697-5.
 76. Park S, Lee S, Kim Y *et al.* Causal effects of relative fat, protein, and carbohydrate intake on chronic kidney disease: a Mendelian randomization study. *Am J Clin Nutr* 2021;113:1023–31.
 77. Li J, Guasch-Ferré M, Chung W *et al.* The Mediterranean diet, plasma metabolome, and cardiovascular disease risk. *Eur Heart J* 2020;41:2645–56.
 78. Park S, Lee S, Kim Y *et al.* Observational or genetically predicted higher vegetable intake and kidney function impairment: an integrated population-scale cross-sectional analysis and Mendelian randomization study. *J Nutr* 2021;nxaa452. doi:10.1093/jn/nxaa452.