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.

Meda R Sandu [1,3 §*], Rhona A Beynon [,2§], Rebecca C Richmond [1,2], Diana L Santos Ferreira [1,2], Lucy Hackshaw-McGeagh [1,3]George Davey Smith[1,2], Chris Metcalfe[1,5], The PRACTICAL Consortium[4], J Athene Lane [1,5, ¤], Richard M Martin [,2, 3¤]

[1] Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, United Kingdom.

[2] Medical Research Council Integrative Epidemiology Unit at the University of Bristol, Bristol, United Kingdom.

[3] The National Institute for Health Research (NIHR) Bristol Biomedical Research Centre, University Hospital Bristol NHS Foundation Trust and University of Bristol, Bristol, United Kingdom.

[4] Members of the PRACTICAL Consortium.

[5] Bristol Randomised Trials Collaboration, Bristol Medical School, University of Bristol, Bristol, United Kingdom.

[§] Authors with equal contributions.

[¤] Co-principle investigators.

[*] Corresponding author

Corresponding author: Meda R Sandu, Department of Population Health Sciences, Bristol Medical School, University of Bristol, NIHR Bristol Biomedical Research Centre Nutrition Theme, 3rd Floor, Education & Research Centre, Upper Maudlin Street, Bristol BS2 8AE, UK Tel: +44 (0)117 3421754, Fax: +44 (0)117 3420239 Email: meda.sandu@bristol.ac.uk

Competing interests: *All authors have completed the Unified Competing Interest form (available on request from the corresponding author) and declare: no support from any organisation for the submitted work; no financial relationships with any organisations that might have an interest in the submitted work in the previous three years, no other relationships or activities that could appear to have influenced the submitted work.*

Abstract

Background

Feasibility trials are preliminary trials that assess the viability and acceptability of intervention studies and the effects of the intervention on intermediate endpoints. Due to their short duration, they are unable to establish the effects of the intervention on long-term clinical outcomes. We propose a novel method that could transform the interpretation of feasibility trials using modified two-stage randomisation analyses.

Methods

In this two-stage process, we explored the effects of a 6-month feasibility factorial randomised controlled trial (RCT) of lycopene and green tea dietary interventions (ProDiet) on 159 serum metabolic traits in 133 men with raised PSA levels but prostate cancer (PCA) free. In the first stage, we conducted an intention-to-treat analysis, using linear regression to examine the effects of the interventions on metabolic traits, compared to the placebo group and instrumental variable analysis to assess the causal effect of the intervention on the outcomes. In the second stage, we used a two-sample Mendelian Randomization (MR) approach to assess the causal effect of metabolic traits altered by the interventions, on PCA risk, using summary statistics data from an international PCA consortium of 44,825 cancer cases and 27,904 controls.

Results

The systemic effects of lycopene and green tea supplementation on serum metabolic profile were comparable to the effects of the respective dietary advice interventions (R^2= 0.65 and 0.76 for lycopene and green tea respectively). Metabolites which were altered in response to lycopene supplementation were acetate (standard deviation difference versus placebo (β)): 0.69; 95% CI= 0.24, 1.15; p=0.003), valine (β: -0.62; -1.03, -0.02; p=0.004), pyruvate (β: -0.56; -0.95, -0.16; p=0.006), and docosahexaenoic acid (β: -0.50; -085, -0.14; p=0.006). The instrumental variable analysis showed there was no evidence that green tea altered the metabolome, but lycopene was associated with an increase in acetate (β=2.13; p=0.006) and decreases in pyruvate (β=-1.90; p=0.009), valine (β=-1.79; p=0.023), diacylglycerol (β=-1.81; p=0.026), alanine (β=-1.55; p=0.015) and DHA (p=0.097), where the regression coefficient represents the standard deviation (SD) difference in metabolite measures per unit change in lycopene (µmol/L) or EGCG (nM).

Using MR, a genetically instrumented SD increase in pyruvate increased the odds of PCA by 1.29 (1.03, 1.62; p=0.027).

Conclusion

Using a two-stage randomization analysis in a feasibility RCT, we found that lycopene lowered levels of pyruvate, which our Mendelian randomization analysis suggests may be causally related to reduced PCA risk.

We propose a new two-step randomisation analysis to enhance the value of feasibility studies of novel interventions, with step one being a conventional feasibility randomised controlled trial to assess the effect of the intervention on short-term intermediate biomarkers and step two being Mendelian randomisation analysis using large-scale genome-wide association studies (GWAS) to appraise the causal effect of any intermediates altered by the intervention on long-term clinical outcomes .

Introduction

Feasibility studies are small-scale studies, including those with a randomised controlled trial (RCT) design, which aim to assess whether and how a novel intervention can be practically tested in a definitive RCT and ultimately implemented in a clinical setting (2). Feasibility trials are not powered to evaluate the effects of the intervention on clinical outcomes and participants are not typically followed-up for enough time to assess long-term outcomes (2). Nonetheless, intermediate endpoints on the causal pathway to clinical outcomes may be measured in feasibility studies in an attempt to predict the clinical and long-term impact of a novel intervention. In this paper we propose a method of robustly predicting the effects of novel interventions on long-term clinical outcomes using data on intermediate endpoints from small feasibility studies.

Box 1. List of Definitions

Feasibility studies are small scale studies that aim to assess whether a large study can be delivered.

Mendelian Randomisation (MR) is an approach which uses genetic variants reliably associated with a potentially modifiable exposure to assess whether associations between modifiable risk factors and disease are likely to be causal.

Genetic variant is an alteration in the DNA sequence, such as a single nucleotide polymorphism (SNP).

Intermediate endpoint is a measurable variable, often a biomarker, which may be on the causal pathway to clinical outcomes.

Biomarker is a biological molecule found in blood, other body fluids, or tissues that is a sign of a normal or abnormal process, or of a condition or disease (1).

Box 2. Key Messages

- Feasibility studies are small scale studies that aim to assess whether a study can be practically implemented and to inform the design of larger trials, but due to low statistical power and short duration of follow-up, feasibility studies cannot assess the clinical outcomes of interest, such as disease risk or progression.
- Mendelian Randomisation is an approach which uses genetic variants reliably associated with a potentially modifiable exposure to assess whether associations between modifiable risk factors and disease are likely to be causal
- Combining feasibility randomised control trials (RCTs) and MR provides randomised evidence of the effect of novel interventions on intermediate endpoints and then, the effect of those intermediates that are altered by the intervention on long-term clinical outcomes.

Challenges of feasibility trials

Feasibility trials aim to assess whether an intervention can be delivered, and a trial performed, in a particular setting. They are typically small in scale and are conducted to

inform the design of larger, definitive trials. Feasibility studies may be implemented to address up to eight areas (Figure 1) (2). However, feasibility studies do not set out, and are not typically powered, to assess the impact of the intervention on the clinical outcomes of interest, such as disease risk or progression. Instead, they can give some indication of the effects of a novel intervention on intermediate short-term endpoints, such as biomarkers of a biological effect of the intervention. These may be established measures of clinical significance, for example prostate specific antigen (PSA) measurements in a study of prostate cancer patients, but also increasingly more novel biomarkers (2, 3) such as measures of the epigenome, metabolome or proteome.

Figure 1. The eight areas of focus of feasibility studies (2)
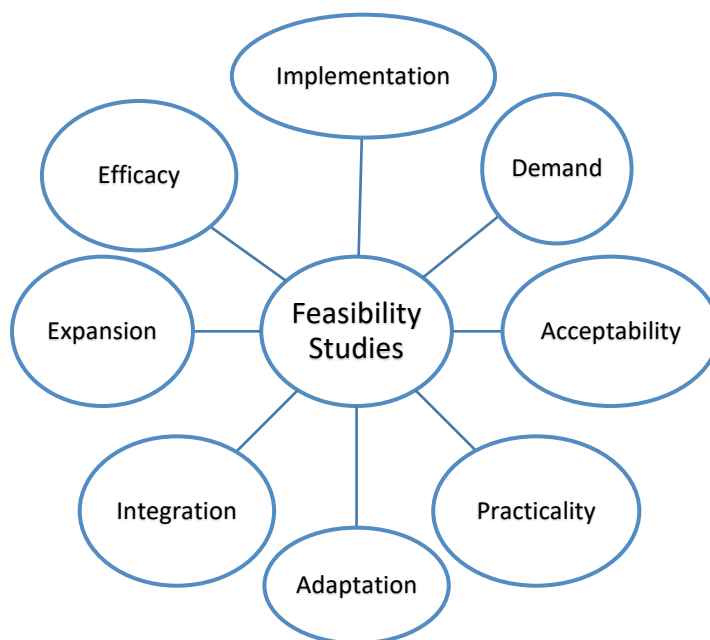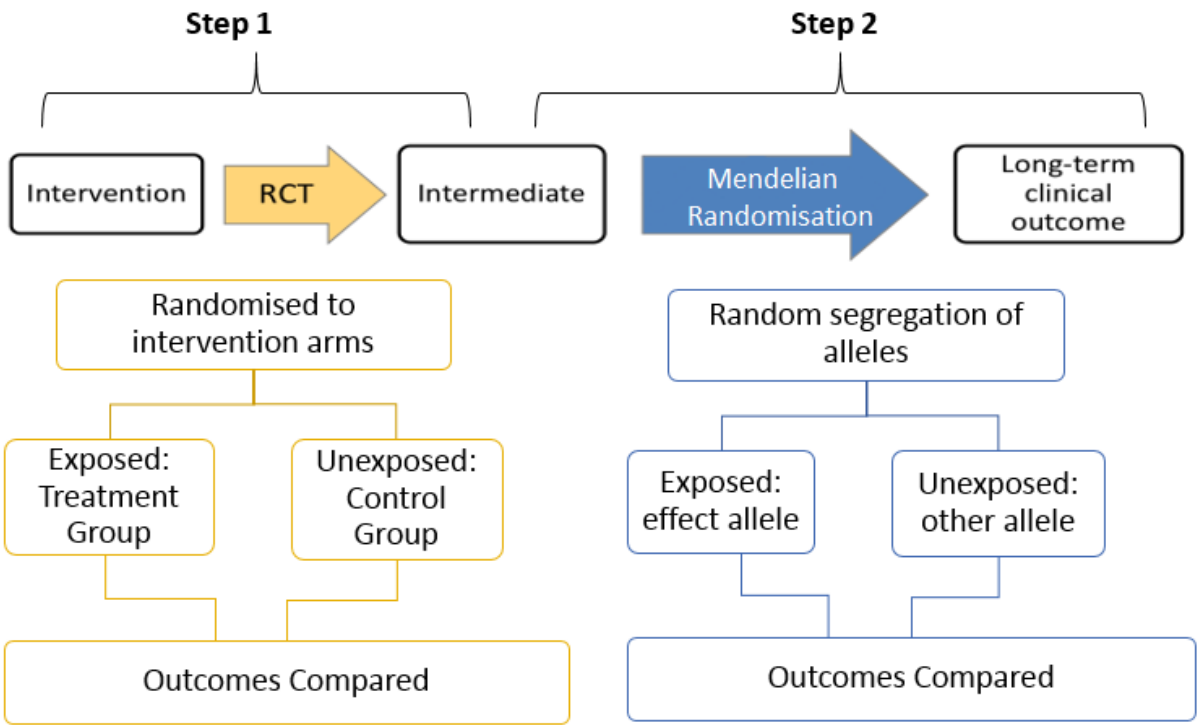

Two-step randomisation in feasibility trials

In a randomised controlled trial (RCT) participants are randomly assigned to multiple groups. Typically, there is an "exposed" group in which participants are randomised to receive an intervention or treatment, and an "unexposed" group in which participants are randomised to not receive the intervention or treatment, commonly referred to as the placebo or control group. In the most robust analysis of RCTs, intention to treat (ITT), the aim is to compare differences between the intervention groups in terms of predefined outcomes (4). ITT however, does not reflect the actual causal effect of the intervention on the outcome, as it does not take into account participant habits (eg non-compliers who do not fully adhere to the intervention). We therefore need to perform an

instrumental variable (IV) analysis, where the assignment to treatment serves as 'instrument' and allows to estimate the causal effect of the intervention on the outcomes.

Mendelian Randomisation (MR) is an analytical technique that uses germline genetic variants robustly associated with an exposure of interest in a natural experiment ("nature's randomised controlled trial") to determine the causal effect of a potentially modifiable exposure on an outcome (5). At conception, germline variants that are associated with a particular trait are randomly distributed with respect to other unrelated traits and environmental factors, akin to random assignment of exposure to treatment or intervention in a conventional RCT (2, 5, 6). Although there are many considerations that should be taken into account when using MR, we can construct an experiment in which people are naturally randomly allocated to a particular exposure based on single nucleotide polymorphisms (SNPs) associated with the exposure of interest(7). Similar to a traditional RCT, and in its simplest form, MR based on one SNP yields an "exposed group" of individuals with a particular genetic variant (allele) which can be compared with an "unexposed group" of individuals without that allele to determine if the outcome occurs more or less frequently in the exposed group compared to the unexposed one. By comparing the outcomes by groups we can establish whether the relationship between exposure and outcome of interest is likely to be causal (8). DNA, although itself unmodifiable, operates through modifiable pathways. MR exploits this to identify modifiable exposures that can be used for disease prevention and therapeutic strategies.

We propose integrating MR into feasibility trials in a two-step process (**Figure 2**). In the first step, ITT and IV analysis are conducted to establish the effects of the novel intervention on intermediate endpoints. Those intermediates altered by the intervention are then investigated further in the second step of the process, an MR analysis. The second step requires the establishment of SNPs that are robustly associated with the intermediates of interest and which can act as genetic instruments for these intermediates. Such genetic instruments are usually identified from previously conducted, large-scale and replicated genome-wide association studies (GWAS). After establishing the SNPs that will be used as instruments, those SNPs are integrated with data from large GWAS consortia investigating the clinical outcome of interest. This second step assesses the causal effect of the intermediates on the outcomes of interest, such as disease risk or progression in a formal MR study. By considering the results from the two steps, it is possible to predict the potential impact of the novel intervention on long-term clinical outcomes, as well as potential adverse effects, via the intermediates.

Figure 2. Two-step randomisation diagram

Worked Example (Pro-Diet RCT)

To illustrate our two-step randomisation approach, we provide a real-life example whereby the results of a small feasibility RCT of dietary interventions for the prevention of prostate cancer were carried forward to a large-scale MR analysis, to infer the causal effect of the interventions on prostate cancer risk via intermediate metabolites (9). The analysis involved two sequential steps (**Figure 2**). Step 1 assessed the randomised effects of lycopene and green tea dietary interventions on 159 serum metabolic traits, quantified by Nuclear Magnetic Resonance (NMR), amongst 132 men enrolled in the ProDiet RCT (10). Step 2 assessed the effects of those metabolic traits altered by the intervention, on prostate cancer risk, using GWAS summary statistics from 44,825 men with prostate cancer and 27,904 controls using publicly available data from the Prostate Cancer Association Group to Investigate Cancer Associated Alterations in the Genome (PRACTICAL) consortium(8, 11, 12).

Figure 3. Two-step randomisation in practice: the ProDiet Randomised Controlled Trial[1] (9)
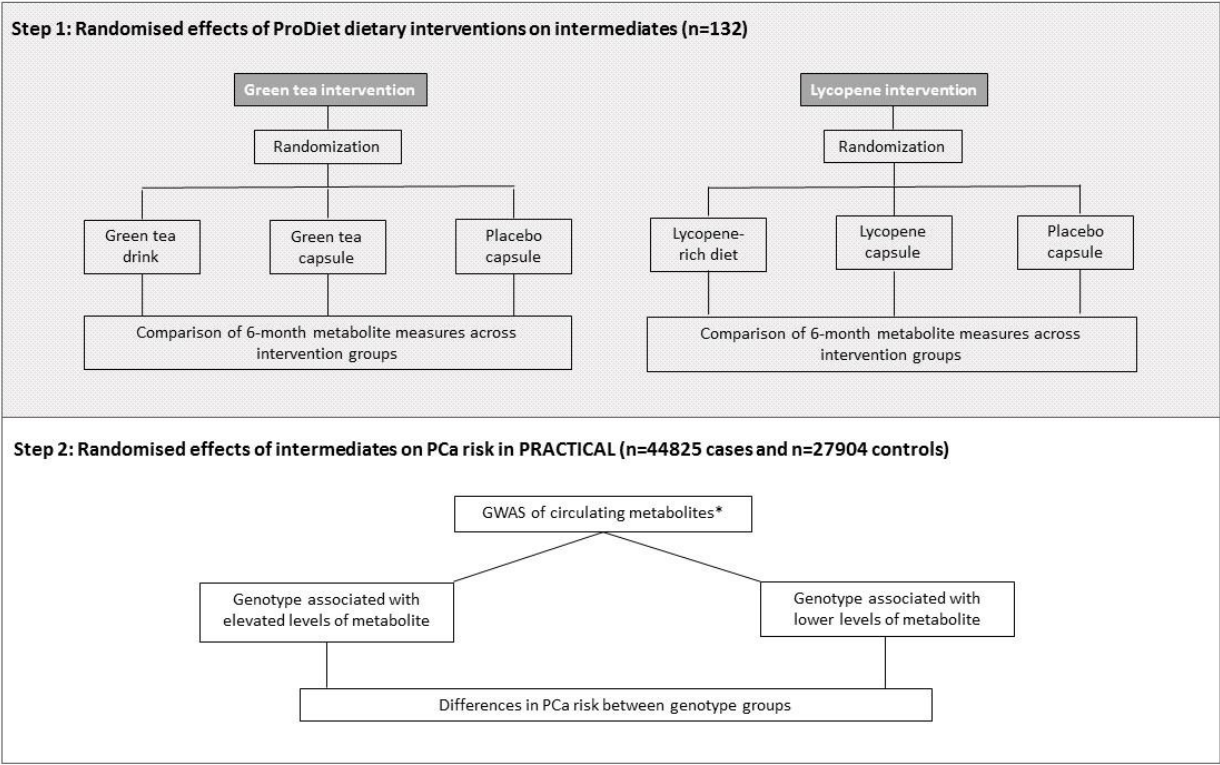
* We used data from Kettunen et al (13) to obtain SNP-associations with metabolic measures.

The ProDiet factorial feasibility RCT (10)

The ProDiet study was approved by the UK Healthcare Research Authority Trent Multicentre Research Ethics Committee (08/H0405/61), conducted according to Declaration of Helsinki, 1964 and the trial number is ISCTRN 95931417(10).The study assessed the feasibility and acceptability of dietary modification in men at increased risk of prostate cancer, with a prostate specific antigen (PSA) level of 2.0-2.95 ng/mL or 3.0-19.95 ng/mL and a negative prostate biopsy (10). Participants were randomised using a 3x3 factorial design, to both a daily green tea (drink at least 3 cups of green tea per day (unblinded) or taking capsules of the active ingredient of green tea (600 mg epigallocatechin-3-gallate (EGCG), (blinded), or placebo (blinded)) and a lycopene intervention (consuming lycopene-rich foods (unblinded) or taking lycopene capsules (15 mg, blinded) or placebo (blinded)), for 6 months (10). Primary endpoints were randomisation rates and intervention adherence (blinded assessment of lycopene and green tea catechin metabolites) at six months. Secondary endpoints included measures of acceptability (from interviews), safety, weight, blood pressure and PSA. Of the 469 men approached, 133 (28.4%) agreed to be randomized and 132 were followed-up (99.2%) (10). The mean serum lycopene was 1.28 (95% confidence intervals (CI) 1.09,

---

[1] PCa: Prostate Cancer

1.50, p = 0.003) times higher in the lycopene supplement group and 1.42 (95% CI 1.21-1.66, p<0.001) times higher in the lycopene-enriched diet group compared to placebo capsule group(10). The median EGCG was 10.7 nM (95% CI 7.0, 32.0) higher in in the active supplement group and 20.0 nM (95% CI 0.0, 19.0) higher in the green tea drink group compared with the placebo capsule group (both p<0.001)(10). The two-step randomisation approach in the ProDiet trial is presented in **Figure 1**.

Step 1: randomised effects of lycopene and green tea on metabolite measures.

In an ITT analysis, linear regression was used to compare standardised metabolic trait measures 6-months after factorial randomisation to lycopene and green tea dietary intervention arms, treating placebo as the reference category (i.e. lycopene supplement or dietary advice versus placebo and green tea supplement or dietary advice versus placebo). The magnitude of the effect of the intervention on circulating metabolic measures was expressed as a standardised mean difference and a new threshold for multiple testing (p-value) was determined using a principle component analysis (PCA) approach, to take account the correlation between metabolic traits. The findings for the ITT in step one of the analysis are presented in **Table 1** and provide some evidence of a change in metabolic measures at follow-up in the lycopene intervention groups, including changes in circulating levels of valine, acetate, pyruvate, diacylglycerol and docosahexaenoic acid (DHA). There was weak evidence to suggest that levels of glycine were lower in the green-tea drinking group compared to placebo.

Using the ITT approach, it was possible to obtain evidence of a randomisation effect, i.e. evidence of a difference in metabolic trait levels between active lycopene/green tea intervention arms and placebo at follow-up, however, it did not provide an estimate of the causal effect of lycopene/EGCG dose on metabolites because some participants may not have taken the capsule/followed the dietary advice, as instructed (i.e. they were non-compliant). IV methods were therefore used to estimate the effect of the lycopene and EGCG dose on the metabolic traits. In this case, the RCT was treated as an IV, with treatment assignment as the "instrument." Using a 2-stage least squares regression approach, the intervention arm was first regressed on the follow-up serum lycopene/EGCG levels. The F-statistics and R^2 values from the first-stage regression were examined to check the instrumental variable assumption that the instrument was sufficiently associated with the exposure. Causal estimates for the instrumented effect of serum lycopene/EGCG levels on each follow-up metabolite were then obtained from the second-stage regression. Results of the IV analysis are presented in Table 2. Again, there was no evidence that green tea altered the metabolome, but lycopene was associated with an increase in acetate (β=2.13; p=0.006) and

decreases in pyruvate (β=-1.90; p=0.009), valine (β=-1.79; p=0.023), diacylglycerol (β=-1.81; p=0.026) and DHA (p=0.097). Alanine was also lower (β=-1.55; p=0.015), where the regression coefficient represents the standard deviation (SD) difference in metabolite measures per unit change in lycopene (μmol/L) or EGCG (nM). These metabolites were then carried forward to the second step of the process.

Table 1: The effects of lycopene and green tea dietary interventions on NMR measured circulating metabolites in ProDiet (Step 1 Results) and the effects of metabolites on prostate cancer risk(PRACTICAL) (step 2)(9, 12, 14).

Step 1 Results: Effects of the interventions on metabolites (ITT)

Metabolite	Intervention arm	Mean difference*	Lower CI	Upper CI	p-value
Lycopene arm					
Valine	Supplement	-0.62	-1.03	-0.2	0.004‡
	Dietary advice	-0.65	-1.04	-0.26	0.001‡
Acetate	Supplement	0.69	0.24	1.15	0.003‡
	Dietary advice	0.26	-0.08	0.59	0.129
Pyruvate	Supplement	-0.56	-0.95	-0.16	0.006
	Dietary advice	-0.3	-0.75	0.15	0.196
Diacylglycerol	Supplement	-0.47	-0.9	-0.03	0.036
	Dietary advice	-0.59	-1.01	-0.18	0.006
DHA	Supplement	-0.5	-0.85	-0.14	0.006
	Dietary advice	-0.15	-0.62	0.32	0.537
Green Tea arm					
Glycine	Supplement	-0.32	-0.79	0.14	0.172
	Dietary advice	-0.58	-0.98	-0.18	0.005
PUFA: FA	Supplement	0.66	0.27	1.05	0.001‡
	Dietary advice	0.43	-0.01	0.86	0.057
Cholesterol esters in small HDL	Supplement	0.22	-0.24	0.67	0.347
	Dietary advice	0.62	0.19	1.04	0.005
Omega-6: FA	Supplement	0.32	-0.12	0.76	0.148
	Dietary advice	0.22	-0.24	0.67	0.005

*Standardized mean difference (and 95% confidence interval [CI] in metabolite trait concentration. Metabolite measures which reached the Bonferroni corrected threshold for multiple testing (P=0.004) are denoted ‡. Abbreviations: DHA, docosahexaenoic acid; FA, fatty acid; HDL, high density lipoprotein; Omega-6:

FA, omega-6 as a proportion of total FA; PUFA: FA, polyunsaturated fatty acids as a proportion of total FA. Omega-6: FA and PUFA: FA are expressed as a % of total FA. Adapted from Beynon *et al* (2018).

Table 2: Lycopene and green tea IV regression results for selected metabolic traits.

Metabolite	N	SD difference	SE	p-value
Lycopene				
Acetate	116	2.13	0.78	0.006
Pyruvate	116	-1.90	0.73	0.009
Alanine	116	-1.55	0.65	0.016
Valine	116	-1.79	0.78	0.023
Diacylglycerol	108	-1.81	0.81	0.026
DHA	116	-0.73	0.44	0.097
Green tea				
Glycine	116	-2.83	3.13	0.366

Results show the standard deviation (SD) difference in metabolite measures per unit change in lycopene (μmol/L) or EGCG (nM). Results are adjusted for baseline metabolic traits concentration. Lycopene: F=2.21; R^2=0.100. Green tea: F=0.31; R^2=0.00. Abbreviations: DHA, docosahexaenoic acid; N, number; SD, standard deviation; SE, standard error. Adapted from Beynon *et al.*, 2018.

Step 2: randomised effects of intermediate metabolites on prostate cancer risk

To appraise the causal role of altered metabolites in prostate cancer risk, SNPs robustly associated with the metabolites of interest at genome-wide significance ($P < 5 \times 10^{-8}$) for the effect of each SNP on the metabolite) were identified from a published GWAS of circulating metabolites[13]. Summary statistics were available for 17 SNPs associated with acetate (n=1 SNP), valine (n=4), pyruvate (n=2), DHA (n=4), and glycine (n=6). SNPs robustly associated with diacylglycerol were not available in the metabolite GWAS summary statistics, and therefore this metabolite could not be instrumented for MR analysis.

Summary statistics from the GWAS of metabolites were used in tandem with summary statistics for 44,825 prostate cancer cases and 27,904 controls in the PRACTICAL consortium (http://practical.icr.ac.uk/) in a two-sample MR analysis[11]. This enabled the calculation of the log odds ratio for prostate cancer per SD increase in metabolic trait levels. MR estimates for each SNP were calculated using the Wald ratio method[15, 16], which calculates the influence of the SNP-outcome effect divided by the SNP-exposure effect. Where multiple SNPs were available, their causal effects were combined together in a fixed-effects inverse-weighted meta-analysis (IVW)

approach. Results of the MR analysis (Table 3) provide some evidence that genetically raised pyruvate increased the odds of prostate cancer (OR per SD increase: 1.29; 95% CI: 1.03, 1.67; p=0.027). Two sensitivity analyses were conducted to assess the robustness of the results to horizontal pleiotropy, where a genetic variant modifies the outcome of interest through a pathway independent of the exposure, which can lead to biased causal estimates: a weighted median approach and MR-Egger regression(12, 17, 18). The weighted median approach can provide reliable causal estimates so long as at least half of the weight in the analysis comes from variants that are valid instruments (i.e. not pleiotropic), whilst MR-Egger allows all variants to have pleiotropic effects, providing they are not proportional to the variants' effects on the exposure [i.e. the so-called Instrument Strength Independent of Direct Effect (InSIDE) assumption].

Table 3. Causal effect estimates of metabolites on prostate cancer in the PRACTICAL consortium

Step 2 Results: Effects of metabolites on prostate cancer risk

Number of SNPs	OR†	95% CI	P-value
4	1.03	0.90, 1.18	0.647
1	0.89	0.63, 1.25	0.501
2	1.29	1.03, 1.62	0.027
4	0.97	0.85, 1.01	0.647
6	0.99	0.92, 1.06	0.787

MR estimates of odds ratios† [OR] (and associated 95% confidence intervals [CI]) of prostate cancer risk per 1 standard deviation [SD] increase in genetically instrumented metabolite levels. Results obtained using the inverse-variance weighted (IVW) method. From Beynon et al (2018).

Conclusions using Two-step Randomisation in the Pro-Diet feasibility RCT

An intervention to increase lycopene intake altered some circulating metabolites amongst men at increased risk of prostate cancer. In our lycopene intervention, pyruvate levels were decreased by 0.56 SD per unit increase in circulating lycopene, a change which our MR analysis suggests is associated with a 15% decrease in prostate cancer risk. This supports previously proposed metabolic pathway mechanisms for prostate cancer cells, where cells are unable to metabolise glucose rapidly and have to resort to an exogenous energy source, such as extracellular pyruvate, which may play a crucial role in tumour development(19, 20).

By combining the results of a small feasibility study with MR, it has been possible to identify a potential intermediate mechanism through which lycopene intake may be influencing prostate cancer risk. These findings may inform the design of larger prostate cancer prevention trials.

Conclusions

Combining feasibility RCTs and MR provides evidence of the causal effect of novel interventions on intermediate endpoints and then, the causal effect of those intermediates that are altered by the intervention on long-term clinical outcomes. This technique could transform the use of feasibility trials, by predicting effects of novel interventions, generating hypotheses on potential mechanisms of disease risk and progression, informing designs of larger definitive trials and providing further evidence on disease aetiology.

Limitations/ Further work

There are limitations associated with this technique. One of the main limitations is the nature of feasibility trials, which cannot be designed to test effects of intervention on all intermediate biomarker levels. We cannot therefore discard other potential effects of the interventions which are not observed in the feasibility trial due to low power, that we may otherwise observe in a definitive, adequately powered trial(2). Also important in this case is the 4th IV assumption, where the observed effect of the intervention or treatment is a local average, that is, it may only apply to a specific subgroup of patients, and it may incorrectly translate at individual level or in a different subgroup, an issue that affects both RCTs and MR studies and may be seen in some clinical settings (21-24). Using data from large GWAS studies, an adequately powered MR can be conducted, which provides confidence in the MR analysis results(25). Unlike observational data, SNPs are not influenced by environmental factors or reverse causation. Nevertheless, using a limited number of genetic instruments, such as in the case of pyruvate, we cannot exclude the possibility of horizontal pleiotropy, a problem in conducting MR which occurs when the SNPs affect the outcome of interest through other pathways other than the risk factor under investigation (6).

The technique presented in this paper is only able to evaluate the causal relationship between the intervention, intermediate endpoint and outcome. If there is no strong evidence of an effect of the intervention on the outcome via the intermediate, the intermediate should not be discarded from future use if it is highly predictive of the outcome, since it is not necessary for it to causally influence the outcome, such as in the case of prediction models. In addition, a lack of evidence of a causal effect of the intermediate endpoint on the outcome also does not suggest that the intervention has

no effect on the clinical outcome, rather it shows there is no causal effect via the presented pathway.

There are various peer-reviewed publications on the use of MR in epidemiology which provide more theoretical and methodological support with regards to conduction and understanding of MR (6, 8, 26). Assessing the pathways through which the SNPs are likely to act and giving careful consideration on the methods employed are vital in order to correctly apply and interpret MR studies(27, 28). Issues such as having multiple biomarkers on separate pathways or multiple biomarkers on the same pathway, incorrectly assigning, defining and quantifying time-dependent exposures should be rigorously addressed and caution should be taken when interpreting and generalising findings(7, 27, 28). More GWAS studies in the 'omics field would allow the identification of a wider variety of biomarkers and provide more robust associations which would increase the number and the reliability of using the two-step randomisation approach in feasibility RCTs. This technique could also be employed in other areas such as behavioural studies. For instance, we can use intermediate biomarkers from a smoking cessation intervention and use MR to assess the causal effect of these biomarkers on pertinent long-term clinical outcomes, such as lung cancer development or mortality. The same technique could also be used in drug development by validating putative biomarker drug targets.

Acknowledgments

References:

1.   National Cancer Institute NCI Dictionary of Cancer Terms [Available from: https://www.cancer.gov/publications/dictionaries/cancer-terms.

2.   Bowen DJ, Kreuter M, Spring B, Cofta-Woerpel L, Linnan L, Weiner D, et al. How we design feasibility studies. Am J Prev Med. 2009;36(5):452-7.

3.   Hackshaw-McGeagh L, Lane JA, Persad R, Gillatt D, Holly JMP, Koupparis A, et al. Prostate cancer - evidence of exercise and nutrition trial (PrEvENT): study protocol for a randomised controlled feasibility trial. Trials. 2016;17(1):123.

4.   Torgerson DJ, Torgerson C. Designing randomised trials in health, education and the social sciences : an introduction. Basingstoke [England] ;: Palgrave Macmillan; 2008.

5.   Smith GD, Ebrahim S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? International journal of epidemiology. 2003;32(1):1-22.

6.   Davies NM, Holmes MV, Davey Smith G. Reading Mendelian randomisation studies: a guide, glossary, and checklist for clinicians. BMJ. 2018;362.

7.   Swanson SA, Tiemeier H, Ikram MA, Hernán MA. Nature as a Trialist?: Deconstructing the Analogy Between Mendelian Randomization and Randomized Trials. Epidemiology. 2017;28(5):653-9.

8.   Smith GD, Ebrahim S. What can mendelian randomisation tell us about modifiable behavioural and environmental exposures? BMJ. 2005;330:1076-9.

9.   Beynon R, Richmond RC, Santos Ferreira DL, Ness AR, May M, Davey Smith G, et al. Investigating the effects of lycopene and green tea on the metabolome of men at risk of prostate cancer: The ProDiet randomised controlled trial. International Journal of Cancer. 2018;0(ja).

10.   Lane JA, Er V, Avery KNL, Horwood J, Cantwell M, Caro GP, et al. ProDiet: A Phase II Randomized Placebo-controlled Trial of Green Tea Catechins and Lycopene in Men at Increased Risk of Prostate Cancer. Cancer Prevention Research. 2018;11(11):687.

11.   The Institute of Cancer Research. Welcome to PRACTICAL London: The Institute of Cancer Research. (Available from http://practical.icr.ac.uk/blog/.).

12.   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. International journal of epidemiology. 2016;45(6):1961-74.

13.   Kettunen J, Tukiainen T, Sarin A-P, Ortega-Alonso A, Tikkanen E, Lyytikäinen L-P, et al. Genome-wide association study identifies multiple loci influencing human serum metabolite levels. Nature Genetics. 2012;44:269.

14.   Schumacher FR, Al Olama AA, Berndt SI, Benlloch S, Ahmed M, Saunders EJ, et al. Association analyses of more than 140,000 men identify 63 new prostate cancer susceptibility loci. Nature Genetics. 2018;50(7):928-36.

15.   Wald A. The Fitting of Straight Lines if Both Variables are Subject to Error. The Annals of Mathematical Statistics. 1940;11(3):284-300.

16.   Wald A. The Fitting of Straight Lines if Both Variables are Subject to Error. Ann Math Statist. 1940;11(3):284-300.

17.   Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. International Journal of Epidemiology. 2015;44(2):512-25.

18.   Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator. Genetic epidemiology. 2016;40(4):304-14.

19.   Gomez-Cebrian N, Rojas-Benedicto A, Albors-Vaquer A, Lopez-Guerrero JA, Pineda-Lucena A, Puchades-Carrasco L. Metabolomics Contributions to the Discovery of Prostate Cancer Biomarkers. Metabolites. 2019;9(3).

20.   Diers Anne R, Broniowska Katarzyna A, Chang C-F, Hogg N. Pyruvate fuels mitochondrial respiration and proliferation of breast cancer cells: effect of monocarboxylate transporter inhibition. Biochemical Journal. 2012;444:561-71.

21.   Lousdal ML. An introduction to instrumental variable assumptions, validation and estimation. Emerg Themes Epidemiol. 2018;15:1-.

22.   Cortés J, González J, Medina M, Vogler M, Vilaró M, Elmore M, et al. Does evidence support the high expectations placed in precision medicine? A bibliographic review [version 5; peer review: 2 approved, 1 approved with reservations, 3 not approved]. 2019;7(30).

23.   Winkelbeiner S, Leucht S, Kane JM, Homan P. Evaluation of Differences in Individual Treatment Response in Schizophrenia Spectrum Disorders: A Meta-analysisEvaluation of Treatment Response Differences in SchizophreniaEvaluation of Treatment Response Differences in Schizophrenia. JAMA Psychiatry. 2019.

24.   Senn S. Mastering variation: variance components and personalised medicine. Statistics in Medicine. 2016;35(7):966-77.

25.   Lawlor DA. Commentary: Two-sample Mendelian randomization: opportunities and challenges. International journal of epidemiology. 2016;45(3):908-15.

26.   Burgess S, Timpson NJ, Ebrahim S, Davey Smith G. Mendelian randomization: where are we now and where are we going? International Journal of Epidemiology. 2015;44(2):379-88.

27.   Holmes MV, Ala-Korpela M, Smith GD. Mendelian randomization in cardiometabolic disease: challenges in evaluating causality. Nat Rev Cardiol. 2017;14(10):577-90.

28.   Tan VY, Yarmolinsky J, Lawlor DA, Timpson NJ. Letter regarding article, "Associations of obesity and circulating insulin and glucose with breast cancer risk: a Mendelian randomization analysis". International Journal of Epidemiology. 2019.