nature portfolio

Peer Review File

Multi-dimensional evidence from the UK Biobank shows the impact of diet and macronutrient intake on aging

Corresponding Author: Dr Chen Zhu

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
1. The introduction does not give a clear rationale for the aims and objectives of the paper and requires some revisions.
2. The authors used a parametric proportional hazard model based on the Gompertz distribution to generate phenotypic age. What was the rationale for this approach?
3. The description of the HDFE is very unclear and requires further elucidation.
4. The details of the MVMR are vague and unhelpful. How were the SNPs selected from the GWAS for dietary composition and macronutrient intake? More details are required as well as chromosome position, risk allele, risk allele frequency, weights (beta), nearest gene from the GWAS. The information needs to be reported in sufficient detail to allow replication.
5. In the diet GWAS paper by Meddens et al, (2020) referenced by the authors the UK Biobank was used as the discovery sample – so there is a possibility of over inflation of the estimates if the selected SNPs are used for an MVMR in UK Biobank. How did the authors deal with this issue?
6. Details of non-linear MR analysis are not reported in the methods section.
7. The UK Biobank due to its large sample size has the ability to detect very small effect sizes. However statistical significance should not be confused with clinical significance. The authors need to interpret there results with more caution due to this aspect.
8. Did the authors consider correcting for multiple testing in their analysis? Which methods did they consider?
9. The authors appear to introduce the results of additional analyses in the discussion section of the manuscript (lines 174 -178 – the authors mention conducting additional HDFE analyses that were not described in the results section).
10. In the discussion the authors refer to their “phenotypic age” as an epigenetic biomarker of ageing. This is not true as this was based on chronological age and nine biochemical markers only. How do the authors define epigenetic biomarkers of ageing? In my view epigenetic biomarkers are based on DNA methylation.
11. The results should be reported in accordance with the STROBE-MR guidelines
12. References 4 and 6 are the same paper
13. The authors should not report p=0.0000 in tables.
14. The tables are of poor quality and need to be re-organised to improve the readability.

Reviewer #2

(Remarks to the Author)
The paper studies the potential effects of diet on various aging markers. This is a relevant question, given that aging should relate to the diet, eg in animal studies. The paper uses UK Biobank as the main sample and uses two methods - a variant of regression called High-dimensional Fixed Effects and then Mendelian randomisation. They found that carbohydrate an plant-based food intake is related to all aging markers whereas other dietary factors had varying effects on different aging markers.

I think it is generally great to have different statistical methods to try to answer the same question. But here, the questions were a bit different. Namely, separate dietary markers were used for HDFE analysis and MR. This way, it is hard to draw conclusions that would replicate in both HDFE and MR analysis.

I would like to have more explanation, what is the benefit of HDFE analysis? How vulnerable is this to biases, compared to

MR? Why do we care about HDFE analyses if MR is supposedly superior?

The dietary patterns as such are vulnerable to their own self-report biases. Pirastu et al describe dealt with these biases extensively in their UK biobank analyses.
https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1010162

It would be great, if the instruments of Pirastu et al., would be used, I think I would have more faith in dietary genetic markers that have considered potential dietary self-report biases.
Comments on specific parts of the paper:
“Associations of dietary factors with aging” (p. 3)
(p. 3) How related are the different aging markers? Given that they have varying associations with diet factors, this makes me wonder, whether they point to the same outcome. Do we expect the aging markers to correlate?
Similarly, it would be great to have correlations of the various predictors.
“three macronutrient intakes simultaneously” (p. 5)
(p. 5) How were the diet components captured? With relative % or absolute numbers? In any case, these numbers tend to be related - if a person has more carbohydrates, they tend to have less fats or proteins. Is that an issue for the MVMR model?
“semiparametric nonlinear Mendelian Randomization (NLMR) estimation.[37,38] Estimates from NLMR confirmed the beneficial linear effects of carbohydrate intake on aging but detected no significant nonlinear effect of carbohydrate intake (nonlinearity p-value=0.386).” (p. 6)
(p. 6) The NLMR method was recently revised/retracted. Does it influence the results found here?

https://twitter.com/mendel_random/status/1731253281161400762

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)

no further comments

Reviewer #2

(Remarks to the Author)
Dear authors

Thank you for the revision. I appreciate the effort you have made to make this paper clearer to read. Unfortunately, the unfamiliar HDFE (at least for me), is still something that would require clearer explanation.

I do not understand the benefit of HDFE over common linear regression. I could easily add all the covariates to a linear model and model the same association. While there is a lot of them, the sample size is huge so i think the linear model could handle them, no?

Are you using multiple outcomes simultaneously? That would be a step up. But this could also be done in SEM framework.

Are the statistical effects independent of all other predictors & covariates?

I am not statistically versed nor do I have the time to read the HDFE paper in detail to understand how it differs from regular linear model. Since you bring a new method to the table, you have to convince us, non-statisticians that it brings something more to the table.

I would also like to see the same associations modelled with a traditional linear regression to see, whether there is in fact any differnece between HDFE and linear regression.

I think the 477 FEs covariates should also be listed. The best would be to share analysis code, so people wishing to do so, could reproduce the code.

I understand that some of the FEs were also used in MR. If you list them, then clarify, which are in MR and which not.

Version 2:

Reviewer comments:

Reviewer #2

(Remarks to the Author)
Thank you for your efforts and background information on HDFE, as well as the analysis script. The new method is more clearly justified and in general, the paper reads better. The methods are more transparent. I have to say that as I have not used STATA nor this MR package, I am not fully able to tell, what exactly is going on. But experts with those packages can now more easily track how you achieved these results!

The MR script makes it very clear, what covariates are used. These variables do not fully line up with the R4 table, there is education, and townsend index, and less ethnic dummies than in table R4. Could you please fix the table R4 to reflect what was done in the script?

Finally, as I said, I found the additional material supplied to me very useful. I think some of your readers may also have similar questions - could you please add the tables R2, R3, and R4 to the supplement and also link to the github repo in the paper? Thanks!

Response to Reviewer 1's Comments

Before formally addressing your insightful and constructive comments, we would like to thank you for your time and energy devoted to reviewing our manuscript. Your comments and suggestions have undoubtedly helped us improve the manuscript. Below, we address each of your comments in a point-by-point manner. For ease of review, we quoted your comments in *italic* type; our responses are provided in a normal font.

Reviewer #1 (Remarks to the Author):

1. The introduction does not give a clear rationale for the aims and objectives of the paper and requires some revisions.

Response: Thank you for pointing this out. We have revised the introduction and provided a clearer rationale for the aims and objectives of the paper (on page 3):

"Aging is a complex biological process involving molecular changes, cellular senescence, and physiological dysregulation. While chronological aging is inevitable, biological aging varies among individuals due to genetic and environmental factors. Geroscience theory posits that slowing biological aging could prevent diseases, extend healthy lifespan, and reduce healthcare costs. Lifestyle factors such as diet, alcohol consumption, smoking, and physical activity significantly influence aging. Diet, in particular, directly impacts health outcomes, including chronic diseases, serum biomarkers, and DNA methylation, suggesting its potential role in modulating aging.

However, research on specific dietary patterns and their effects on aging has yielded inconsistent and inconclusive results. This is primarily due to diverse methods of measuring biological aging, limited sample sizes, and a lack of controlled trials. For example, while some studies suggest that low-carbohydrate diets improve serum factors related to aging, others link these diets to higher mortality rates. Similarly, reviews on diet's impact on cognitive aging and telomere length report mixed outcomes.

Given the imperative to promote healthy aging—a state where individuals maintain functional abilities that enable well-being—the lack of clear guidance from existing studies is a significant gap in our knowledge. This study aims to systematically investigate the relationships between dietary factors and multidimensional measures of aging (i.e., telomere length, blood-based phenotypical age, grey matter volume, and white matter volume), focusing particularly on the causal impacts of macronutrient intakes using cohort study data. Our goal is to clarify the role of diet in biological aging processes and provide robust evidence to support dietary recommendations for healthy aging."

2. The authors used a parametric proportional hazard model based on the Gompertz distribution to generate phenotypic age. What was the rationale for this approach?

Response: Thank you for pointing this out. We opted to use a parametric proportional hazard model based on the Gompertz distribution to construct phenotypic age, following the methodology established by Yang et al. (2022, *JAMA Network Open*). This approach integrates nine clinical biomarkers and chronological age to model the 10-year mortality risk, converting it into an equivalent "phenotypic age" metric. This method is particularly robust in aging research for its ability to provide a quantifiable measure of biological aging based on mortality risk. We have expanded on this rationale in the revised manuscript to clarify our methodological choices and align with established protocols (on page 6 of the revised manuscript):

"Blood-based phenotypic age was determined using chronological age and nine clinical biomarkers as proposed by Yang et al. (2022): albumin, creatinine, glucose, C-reactive protein, lymphocyte percentage, mean corpuscular volume, red cell distribution width, alkaline phosphatase, and white blood cell count. Employing a parametric proportional hazards model based on the Gompertz distribution, this approach calculates a "phenotypic age" by modeling 10-year mortality risk, offering a robust quantifiable measure of biological aging. Biomarker data were obtained from UK Biobank samples collected at enrollment, typically analyzed within 24 hours at the central laboratory using Beckman Coulter LH750 instruments."

Reference: Yang, G., Cao, X., Li, X., Zhang, J., Ma, C., Zhang, N., ... & Liu, Z. (2022). Association of unhealthy lifestyle and childhood adversity with acceleration of aging among UK biobank participants. *JAMA network open*, *5*(9), e2230690-e2230690.

3. The description of the HDFE is very unclear and requires further elucidation.

Response: Thank you for pointing this out. In our study, we employed the state-of-the-art high-dimensional fixed effects (HDFE) model, developed by Guimaraes and Portugal in 2010, which allows for the inclusion of numerous fixed effects. This methodology is essential for accurately identifying the influence of confounding variables across varied ethnic, regional, and occupational groups. HDFE models surpass traditional multivariate linear regression in precision by effectively mitigating omitted variable bias, thus providing a more reliable and robust analysis of the impacts of key independent variables. We have clarified this approach further in the revised manuscript to ensure better understanding (on page 7).

Reference: Guimaraes, P., & Portugal, P. (2010). A simple feasible procedure to fit models with high-dimensional fixed effects. *The Stata Journal*, *10*(4), 628-649.

4. The details of the MVMR are vague and unhelpful. How were the SNPs selected from the GWAS for dietary composition and macronutrient intake? More details are required as well as

chromosome position, risk allele, risk allele frequency, weights (beta), nearest gene from the GWAS. The information needs to be reported in sufficient detail to allow replication.

Response: Thank you for pointing this out. Multivariable Mendelian Randomization (MVMR) extends traditional Mendelian Randomization (MR) by incorporating multiple genetic instruments for **multiple exposures simultaneously**, thus addressing potential confounding and pleiotropic effects. This provides a more nuanced analysis compared to the regular single-exposure MR. In our study, MVMR was employed to consider the interplay between different macronutrient intakes. We have expanded the description of the MVMR methodology in the revised manuscript for clarity (on page 7 to 8).

The single nucleotide polymorphisms (SNPs) used as genetic instruments for dietary composition and macronutrient intake were selected from GWAS results reported by Meddens et al. (2022) and Pirastu et al. (2022). Detailed information on these SNPs, including chromosome, position, effect allele, alternative allele, effect allele frequency (EAF), beta values, *p* values, and nearest gene, is presented below (Table R1) and has been added to the revised manuscript (Supplementary eTable 3) to ensure transparency and facilitate replication.

Table R1. Detailed information on genetic instrumental variables used in Mendelian Randomization

rsID	CHR	Position	Effect Allele	ALT Allele	EAF	P value	Beta	Nearest gene	Function
rs445551	2	79470856	A	G	0.2452	1.49E-08	0.019	*CTNNA2*	Intron Variant
rs10206338	2	59982846	A	G	0.6063	1.52E-08	-0.016	*AC007100.1*	Intron Variant
rs10510554	3	25058285	T	C	0.4103	2.94E-12	0.019	*AC092422.1*	Intron Variant
rs57193069	7	1822781	A	G	0.5537	1.80E-08	-0.016	*MAD1L1*	Intron Variant
rs1461729	8	9329732	A	G	0.0943	4.09E-12	0.032	*LOC157273*	Intron Variant
rs10962121	9	15702706	T	G	0.3964	3.40E-08	-0.015	*CCDC171*	Intron Variant
rs2472297	15	74735539	T	C	0.2234	3.73E-08	-0.018	*CYP1A1*	Intron Variant
rs8097672	18	1839600	A	T	0.8514	1.54E-12	0.03	*AP005230.1*	Intron Variant
rs33988101	19	48714854	T	G	0.4955	1.66E-26	-0.029	*MAMSTR*	Synonymous Variant

Reference: Meddens, S. F. W., de Vlaming, R., Bowers, P., Burik, C. A., Linnér, R. K., Lee, C., ... & Koellinger, P. D. (2021). Genomic analysis of diet composition finds novel loci and associations with health and lifestyle. *Molecular psychiatry*, *26*(6), 2056-2069.

Pirastu, N., McDonnell, C., Grzeszkowiak, E. J., Mounier, N., Imamura, F., Merino, J., ... & Wilson, J. F. (2022). Using genetic variation to disentangle the complex relationship between food intake and health outcomes. *PLoS Genetics*, *18*(6), e1010162.

Oscanoa, J., Sivapalan, L., Gadaleta, E., Dayem Ullah, A. Z., Lemoine, N. R., & Chelala, C. (2020). SNPnexus: a web server for functional annotation of human genome sequence variation (2020 update). *Nucleic acids research*, *48*(W1), W185-W192.

5. In the diet GWAS paper by Meddens et al, (2020) referenced by the authors the UK Biobank was used as the discovery sample – so there is a possibility of over inflation of the estimates if the selected SNPs are used for an MVMR in UK Biobank. How did the authors deal with this issue?

Response: Thank you for pointing this out. In the study by Meddens et al. (2020), the UK Biobank served as the discovery sample for GWAS, with replication analyses conducted across **14 independent cohorts** from the Netherlands, UK, USA, and various international consortia. To address potential sample overlap, we selected SNPs that achieved genome-wide significance in both the discovery and replication phases, minimizing the risk of inflated estimates.

Moreover, while Burgess et al. (2016) highlighted potential biases in *two-sample* Mendelian randomization due to sample overlap, our study uses a *one-sample* MVMR approach. Burgess et al. confirm that such one-sample approaches are **asymptotically unbiased** (see page 598: *"…estimates from IV analysis in a one-sample setting [that is, genetic variants, risk factor and outcome are all measured in the same participants] are asymptotically unbiased…"*), suggesting that MVMR estimates are not likely to be susceptible to the sample overlap issue.

Reference: Meddens, S. F. W., de Vlaming, R., Bowers, P., Burik, C. A., Linnér, R. K., Lee, C., ... & Koellinger, P. D. (2021). Genomic analysis of diet composition finds novel loci and associations with health and lifestyle. *Molecular psychiatry*, *26*(6), 2056-2069.

Burgess, S., Davies, N. M., & Thompson, S. G. (2016). Bias due to participant overlap in two‐sample Mendelian randomization. *Genetic epidemiology*, *40*(7), 597-608.

6. Details of non-linear MR analysis are not reported in the methods section.

Response: Thank you for your comment regarding the inclusion of non-linear Mendelian Randomization (NLMR) analysis. Initially, we employed NLMR to investigate the potential non-linear effects of macronutrient intake on aging. However, recent critical evaluations, including those by Davey Smith (2023) and Wade et al. (2023), have questioned the reliability of the residual stratification-based NLMR method. In light of these concerns and the retraction of a pivotal paper by Sofianopoulou et al. (2021), we have opted to exclude the NLMR analysis from our revised manuscript. This change was made to maintain the scientific integrity and reliability of our findings, focusing instead on the robust linear effects of macronutrient intake on aging.

Reference: Davey Smith, G. (2023). Mendelian randomisation and vitamin D: the importance of model assumptions. *Lancet Diabetes Endocrinol*, 11(1):14. doi:10.1016/S2213-8587(22)00345-X

Sofianopoulou, E., Kaptoge, S. K., Afzal, S., Jiang, T., Gill, D., Gundersen, T. E., ... & Burgess, S. (2021). RETRACTED: Estimating dose-response relationships for vitamin D with coronary heart disease, stroke, and all-cause mortality: observational and Mendelian randomisation analyses. *The Lancet Diabetes & Endocrinology*, *9*(12), 837-846.

Wade, K. H., Hamilton, F. W., Carslake, D., Sattar, N., Davey Smith, G., & Timpson, N. J. (2023). Challenges in undertaking nonlinear Mendelian randomization. *Obesity*, *31*(12), 2887-2890.

7. The UK Biobank due to its large sample size has the ability to detect very small effect sizes. However statistical significance should not be confused with clinical significance. The authors need to interpret there results with more caution due to this aspect.

Response: Thank you for this important point. We acknowledge the large sample size of the UK Biobank allows for the detection of very small effect sizes, which may not always correspond to clinical significance. In response, we have carefully considered the clinical relevance of our findings. Lemaitre et al. (2012) estimated that the annual grey matter volume (GMV) reduction is 1.89 cm^3 due to aging. Notably, our results show that an increase of 100 grams in daily carbohydrate intake is significantly associated with an increase in grey matter volume (GMV) by 2.62 cm^3 (Table 2, Sidak-Holm corrected p value = 0.040), which potentially offsets brain aging by approximately 1.4 years. Furthermore, our findings suggest that the benefits from this dietary adjustment could be comparable to those achieved through physical activity, as reported in Koblinsky et al. (2021), where physical activity increased total GMV by 2.17 cm^3. We have incorporated these discussions into the revised manuscript (on pages 12 and 13).

References:

Lemaitre, H., Goldman, A. L., Sambataro, F., Verchinski, B. A., Meyer-Lindenberg, A., Weinberger, D. R., & Mattay, V. S. (2012). Normal age-related brain morphometric changes: nonuniformity across cortical thickness, surface area and gray matter volume?. *Neurobiology of aging*, *33*(3), 617-e1.

Koblinsky, N. D., Meusel, L. A. C., Greenwood, C. E., & Anderson, N. D. (2021). Household physical activity is positively associated with gray matter volume in older adults. *BMC geriatrics*, *21*, 1-10.

8. Did the authors consider correcting for multiple testing in their analysis? Which methods did they consider?

Response: Thank you for highlighting this important aspect of our analysis. In response to your query, we have implemented the Sidak-Holm method for multiple hypothesis testing, as outlined by Ludbrook (1998) and Guo et al. (2007), to enhance the statistical validity of our findings. This adjustment led to some originally significant associations, such as the intakes of poultry with shorter telomere lengths and tea with larger white matter volume, becoming insignificant after the correction.

We have revised the Results section based on Sidak-Holm corrected p-values for key estimates, reinforcing the robustness of our conclusions against the risk of type I errors due to multiple

comparisons. This information can be found on pages 9 and 12 , Figure 2, Table 2, and Supplementary eTable 4.

Reference: Ludbrook, J. (1998). Multiple comparison procedures updated. *Clinical and Experimental Pharmacology and Physiology*, *25*(12), 1032-1037.

Guo, W., & Romano, J. (2007). A generalized Sidak-Holm procedure and control of generalized error rates under independence. *Statistical applications in genetics and molecular biology*, *6*(1).

9. The authors appear to introduce the results of additional analyses in the discussion section of the manuscript (lines 174 -178 – the authors mention conducting additional HDFE analyses that were not described in the results section).

Response: Thank you for bringing this to our attention. In our original submission, we aimed to demonstrate the robustness of the main findings concerning caloric intake/calorie restriction but did not adequately structure the manuscript. We have now reorganized the manuscript to improve clarity and coherence. Specifically, we have moved the results of the additional high-dimensional fixed effects (HDFE) analyses, which test the role of caloric intake/restriction, to the results section of the revised manuscript (page 11, subsection "*Calorie restriction and aging*").

10. In the discussion the authors refer to their "phenotypic age" as an epigenetic biomarker of ageing. This is not true as this was based on chronological age and nine biochemical markers only. How do the authors define epigenetic biomarkers of ageing? In my view epigenetic biomarkers are based on DNA methylation.

Response: Thank you for highlighting this issue. We use the term "phenotypic age" following the original work by Yang et al. (2022, *JAMA Network Open*). In their study, the authors integrated nine clinical biomarkers and chronological age to create a "phenotypic age" metric, demonstrating its robustness in aging research. We agree with your concern that "phenotypic age" may be misleading, so we have revised the term to "clinical biomarkers-based phenotypic age" in the manuscript.

Regarding epigenetic biomarkers of aging, both telomere length and DNA methylation are recognized as such in the literature (Pearce et al., 2022; Henriques et al., 2024). In our study, telomere length was used as an epigenetic marker since DNA methylation data was unavailable in the UK Biobank dataset.

Reference: Yang, G., Cao, X., Li, X., Zhang, J., Ma, C., Zhang, N., ... & Liu, Z. (2022). Association of unhealthy lifestyle and childhood adversity with acceleration of aging among UK biobank participants. *JAMA network open*, *5*(9), e2230690-e2230690.

Pearce, E. E., Alsaggaf, R., Katta, S., Dagnall, C., Aubert, G., Hicks, B. D., ... & Gadalla, S. M. (2022). Telomere length and epigenetic clocks as markers of cellular aging: a comparative study. *Geroscience*, *44*(3), 1861-1869.

Henriques, C. M., & Ferreira, M. G. (2024). Telomere length is an epigenetic trait–Implications for the use of telomerase-deficient organisms to model human disease. *Disease Models & Mechanisms*, *17*(3), dmm050581.

11. The results should be reported in accordance with the STROBE-MR guidelines

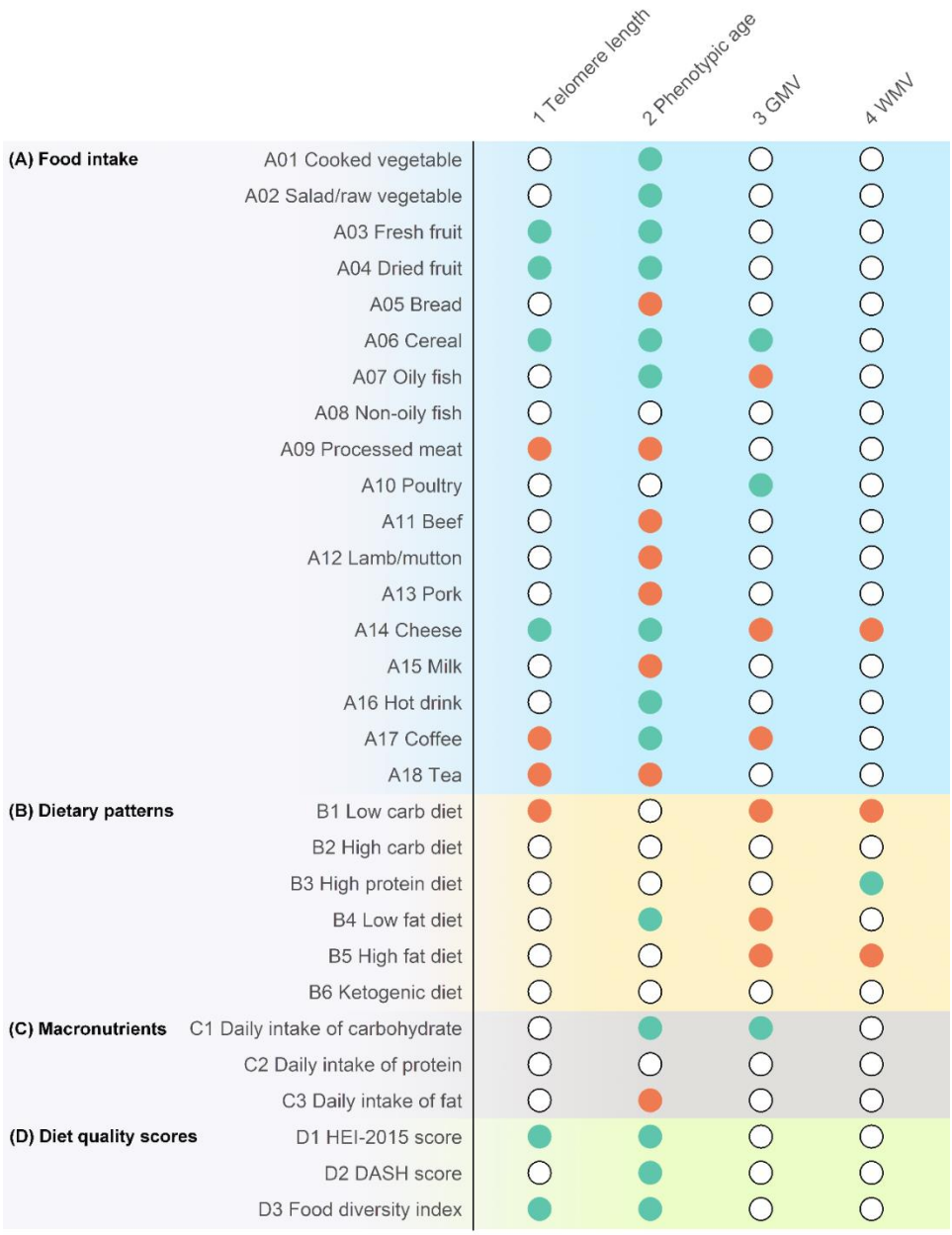
Response: Thanks for reminding us about this. We have revised the manuscript according to the STROBE-MR guidelines, and include a STROBE-MR Checklist in the Supplimentary Materials.

12. References 4 and 6 are the same paper
Response: Thank you for spotting this error. We have dropped reference 6 and renumbered the references.

13. The authors should not report p=0.0000 in tables.
Response: Thank you for pointing this out. We have replaced all p=0.0000 with p<0.0001.

14. The tables are of poor quality and need to be re-organised to improve the readability.
Response: Thank you for pointing this out. We have revised the presentation of our results to enhance readability. Specifically, key estimates from the HDFE regressions are now clearly visualized as a concise heatmap provided. As shown in Figure R1 below, the heatmap summerizes estimated effects of all dietary exposures (in rows) on different aging outcomes (in columns), where red dots indicate potential risky effects, green dots indicate protective effects, and white dots indicate null (i.e., insignificant after the Sidak-Holm correction for multiple testing) effects on specific measures of aging. And we have included it as Figure 2 in the revised manuscript. Additionally, detailed information about beta values and p-values, both original and those corrected for multiple testing, are now thoroughly documented in Supplementary eTable 4. This restructuring aims to make the results more comprehensible.

Figure R1. HDFE estimates of dietary factors on aging outcomes


1 Telomere length
2 Phenotypic age
3 GMV
4 WMV
(A) Food intake
A01 Cooked vegetable
A02 Salad/raw vegetable
A03 Fresh fruit
A04 Dried fruit
A05 Bread
A06 Cereal
A07 Oily fish
A08 Non-oily fish
A09 Processed meat
A10 Poultry
A11 Beef
A12 Lamb/mutton
A13 Pork
A14 Cheese
A15 Milk
A16 Hot drink
A17 Coffee
A18 Tea
(B) Dietary patterns
B1 Low carb diet
B2 High carb diet
B3 High protein diet
B4 Low fat diet
B5 High fat diet
B6 Ketogenic diet
(C) Macronutrients
C1 Daily intake of carbohydrate
C2 Daily intake of protein
C3 Daily intake of fat
(D) Diet quality scores
D1 HEI-2015 score
D2 DASH score
D3 Food diversity index

Response to Reviewer 2's Comments

Before formally addressing your insightful and constructive comments, we would like to thank you for your time and energy devoted to reviewing our manuscript. Your comments and suggestions have undoubtedly helped improve the manuscript. Below, we address each of your comments in a point-by-point manner. For ease of review, we quoted your comments in *italic* type; our responses are provided in a normal font. If we may, we have numbered your comments for ease of addressing.

Reviewer #2 (Remarks to the Author):

1.The paper studies the potential effects of diet on various aging markers. This is a relevant question, given that aging should relate to the diet, eg in animal studies. The paper uses UK Biobank as the main sample and uses two methods - a variant of regression called High-dimensional Fixed Effects and then Mendelian randomisation. They found that carbohydrate an plant-based food intake is related to all aging markers whereas other dietary factors had varying effects on different aging markers.

Response: Thank you for the excellent summary. We address all your other comments below in a point-by-point manner.

2. I think it is generally great to have different statistical methods to try to answer the same question. But here, the questions were a bit different. Namely, separate dietary markers were used for HDFE analysis and MR. This way, it is hard to draw conclusions that would replicate in both HDFE and MR analysis.

Response: Thank you for your insightful comments. In our study, we used the high-dimensional fixed effects (HDFE) model (Guimaraes and Portugal, 2010) as the primary statistical approach. This method effectively includes numerous fixed effects, accurately identifying the influence of confounding variables across different ethnic, regional, and occupational groups. HDFE models are more precise than traditional multivariate linear regression (e.g., OLS) by mitigating omitted variable bias, providing a more robust analysis of key independent variables.

To further strengthen the causal link between dietary factors and aging, we also employed Mendelian randomization (MR). We selected only a subset of dietary markers—specifically, dietary intakes of carbohydrates, proteins, and fats—for MR, based on well-established genetic variants from previous genome-wide association studies (Meddens et al., 2021). This selection was due to the robust data available for these macronutrients, ensuring the reliability of MR results.

The combination of the HDFE approach and MR analysis forms a triangulation design, which strengthens findings when different statistical methods, based on distinct assumptions, reach the same conclusion (Lawlor et al., 2016; Burgess and Thompson, 2021). Therefore, the MR results provide additional triangulation evidence on the causal effects of carbohydrate, protein, and fat

intakes on aging, further supporting the causal link between macronutrient intakes and aging. We have clarified this in the revised manuscript to enhance understanding and transparency (on pages 6 to 9).

Reference:

Guimaraes, P., & Portugal, P. (2010). A simple feasible procedure to fit models with high-dimensional fixed effects. *The Stata Journal*, *10*(4), 628-649.

Meddens, S. F. W., de Vlaming, R., Bowers, P., Burik, C. A., Linnér, R. K., Lee, C., ... & Koellinger, P. D. (2021). Genomic analysis of diet composition finds novel loci and associations with health and lifestyle. *Molecular psychiatry*, *26*(6), 2056-2069.

Lawlor, D. A., Tilling, K., & Davey Smith, G. (2016). Triangulation in aetiological epidemiology. *International journal of epidemiology*, *45*(6), 1866-1886.

Burgess, S., & Thompson, S. G. (2021). *Mendelian randomization: methods for causal inference using genetic variants*. CRC Press.

3. I would like to have more explanation, what is the benefit of HDFE analysis? How vulnerable is this to biases, compared to MR? Why do we care about HDFE analyses if MR is supposedly superior?

Response: Thank you for your insightful comments. The use of High-Dimensional Fixed Effects (HDFE) analysis in our study offers several benefits. HDFE models allow us to control for numerous fixed effects simultaneously, effectively accounting for confounding variables across various groups such as ethnic, regional, and occupational categories. This method mitigates omitted variable bias, providing a robust analysis of key independent variables.

Mendelian Randomization (MR), on the other hand, leverages genetic instrumental variables to derive causal inferences, strengthening findings from HDFE results. However, MR is more stringent compared to HDFE, requiring robust and well-established associations between genetic variants and exposures. This means MR may not always be applicable, especially when no known genetic variants are linked to the exposure. In another words, MR can only be applied to exposures or risk factors for which suitable genetic variants are already identified and available from previous GWAS studies (Larsoon et al., 2023)

By employing both HDFE and MR, we harness the strengths of each method. HDFE offers comprehensive analyses of the associations between a broad range of dietary factors and aging, while MR provides robust causal inference through genetic instruments when applicable. This dual approach ensures our findings are robust and reliable, balancing the strengths and limitations of each method. We have clarified these points in the revised manuscript to provide a clearer understanding of the rationale and benefits of using both HDFE and MR analyses.

Reference: Larsson, S. C., Butterworth, A. S., & Burgess, S. (2023). Mendelian randomization for cardiovascular diseases: principles and applications. *European heart journal*, *44*(47), 4913-4924.

3. The dietary patterns as such are vulnerable to their own self-report biases. Pirastu et al describe dealt with these biases extensively in their UK biobank analyses.

https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1010162
It would be great, if the instruments of Pirastu et al., would be used, I think I would have more faith in dietary genetic markers that have considered potential dietary self-report biases.

Response: Thank you for your insightful comments. In our study, we utilized dietary data from self-reported frequencies of food and beverage consumption similar to those used in Pirastu et al. (2022), derived from the UK Biobank's touchscreen dietary frequency questionnaire. While these data sources are subject to self-report biases, their validity has been supported by studies like Bradbury et al. (2018), which confirmed their reliability in ranking intakes of major food groups. Recognizing the limitations of self-reported data, we have noted this in the Discussion section of our revised manuscript.

Moreover, following your suggestion, we have adopted Pirastu et al. (2022)'s approach and confirmed that only non-mediated genetic variants were selected as instrumental variables for our Mendelian randomization analyses. This adjustment aims to refine the robustness of our causal inferences by reducing bias associated with potential mediation effects. These enhancements are detailed in the revised manuscript (on page 8):

"…To rule out potential mediating effects of genetic variants on aging outcomes via other traits that could violate the statistical assumptions of MR designs (Pirastu et al., 2022), these SNPs were selected based on functional annotation results from SNPnexus and GWAS Catalog online tools (Oscanoa et al., 2020; MacArthur et al., 2017). Detailed information on these SNPs, including chromosome, position, effect allele, alternative allele, effect allele frequency, beta values, p-values, and nearest gene were provided in Supplementary eTable 3 …"

Reference: Pirastu, N., McDonnell, C., Grzeszkowiak, E. J., Mounier, N., Imamura, F., Merino, J., ... & Wilson, J. F. (2022). Using genetic variation to disentangle the complex relationship between food intake and health outcomes. *PLoS Genetics*, *18*(6), e1010162.

Bradbury, K. E., Young, H. J., Guo, W., & Key, T. J. (2018). Dietary assessment in UK Biobank: an evaluation of the performance of the touchscreen dietary questionnaire. *Journal of nutritional science*, *7*, e6.

Oscanoa, J., Sivapalan, L., Gadaleta, E., Dayem Ullah, A. Z., Lemoine, N. R., & Chelala, C. (2020). SNPnexus: a web server for functional annotation of human genome sequence variation (2020 update). *Nucleic acids research*, 48(W1), W185-W192.

MacArthur, J., Bowler, E., Cerezo, M., Gil, L., Hall, P., Hastings, E., ... & Parkinson, H. (2017). The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic acids research*, 45(D1), D896-D901.

4. Comments on specific parts of the paper:

4.1. "Associations of dietary factors with aging" (p. 3)

(p. 3) How related are the different aging markers? Given that they have varying associations with diet factors, this makes me wonder, whether they point to the same outcome. Do we expect the aging markers to correlate?

Response: Thank you for your insightful question. We have included a table below (Table R2) that displays the correlations between the aging markers used in our study. All correlation coefficients are significant at the 1% level, with values ranging from 0.06 to 0.60. Except for blood-based phenotypical age, all pairwise correlation coefficients are positive, indicating that while the aging markers are partially related, they also capture distinct aspects of aging. This reflects the multifaceted nature of the aging process. Additionally, based on your suggestion, we have added heatmaps showing pairwise Pearson's correlation coefficients across outcome variables and dietary exposures in Figure 1, panel a and b of the revised manuscript.

Table R2. Pairwise Pearson's correlation coefficients of four aging markers

Variables	(1) Telomere Length	(2) Biological Age	(3) Grey Matter Volume	(4) White Matter Volume
(1) Telomere Length	1			
(2) Blood-based phenotypical age	-0.20	1		
(3) Grey Matter Volume	0.13	-0.60	1	
(4) White Matter Volume	0.06	-0.23	0.35	1

4.2. Similarly, it would be great to have correlations of the various predictors.

Response: Thank you for the comment. As responded to the previous question, we have added a heatmap showing pairwise correlation coefficients across various dietary predictors in panel b of Figure 1 of the revised manuscript.

4.3. "three macronutrient intakes simultaneously" (p. 5)

(p. 5) How were the diet components captured? With relative % or absolute numbers? In any case, these numbers tend to be related - if a person has more carbohydrates, they tend to have less fats or proteins. Is that an issue for the MVMR model?

Response: Thank you for raising this important point. The daily intake of three macronutrients—carbohydrates, fat, and protein—were measured in absolute values, as described by the UK Biobank (Field IDs: 26013, 26008, and 26005), with average intakes of 254.9 grams, 77.9 grams, and 82.4 grams, respectively (as reported in Table 1 in the original and revised manuscripts).

We agree with your insight on the interrelatedness of these intakes, which is why we employ the Multivariable Mendelian Randomization (MVMR) model. Unlike standard univariable MR, MVMR allows us to *simultaneously* estimate the effects of carbohydrate, fat, and protein intakes, thus maximizing efficiency and reducing bias (Burgess and Thompson,

2015; Sanderson et al., 2019). This approach is analogous to assessing several treatments simultaneously in a factorial randomized trial. We have noted this in the Methods section of our revised manuscript (on pages 7 to 8).

Reference: Burgess, S., & Thompson, S. G. (2015). Multivariable Mendelian randomization: the use of pleiotropic genetic variants to estimate causal effects. American journal of epidemiology, 181(4), 251-260.

Sanderson, E., Davey Smith, G., Windmeijer, F., & Bowden, J. (2019). An examination of multivariable Mendelian randomization in the single-sample and two-sample summary data settings. International journal of epidemiology, 48(3), 713-727.

4.4."semiparametric nonlinear Mendelian Randomization (NLMR) estimation.[37,38] Estimates from NLMR confirmed the beneficial linear effects of carbohydrate intake on aging but detected no significant nonlinear effect of carbohydrate intake (nonlinearity p-value=0.386)." (p. 6)
(p. 6) The NLMR method was recently revised/retracted. Does it influence the results found here?

https://twitter.com/mendel_random/status/1731253281161400762

Response: Thank you for raising this important point. We have also noted recent critical evaluations by Davey Smith (2023) and Wade et al. (2023), questioning the reliability of the residual stratification-based NLMR method. In light of these concerns and the retraction of a pivotal paper by Sofianopoulou et al. (2021), as mentioned in the tweet you referred to, we have decided to exclude the NLMR analysis from our revised manuscript. This decision ensures the scientific integrity and reliability of our findings, focusing on the robust linear effects of macronutrient intake on aging.

Reference: Davey Smith, G. (2023). Mendelian randomisation and vitamin D: the importance of model assumptions. *Lancet Diabetes Endocrinol*, 11(1):14. doi:10.1016/S2213-8587(22)00345-X

Wade, K. H., Hamilton, F. W., Carslake, D., Sattar, N., Davey Smith, G., & Timpson, N. J. (2023). Challenges in undertaking nonlinear Mendelian randomization. *Obesity*, *31*(12), 2887-2890.

Sofianopoulou, E., Kaptoge, S. K., Afzal, S., Jiang, T., Gill, D., Gundersen, T. E., ... & Burgess, S. (2021). RETRACTED: Estimating dose-response relationships for vitamin D with coronary heart disease, stroke, and all-cause mortality: observational and Mendelian randomisation analyses. *The Lancet Diabetes & Endocrinology*, *9*(12), 837-846.

Response to Reviewer Comments (Round 2)

Before formally addressing your insightful and constructive comments, we would like to thank you for your time and energy devoted to reviewing this manuscript. Your comments and suggestions have undoubtedly helped improve the manuscript, again. Below, we address your comments in a point-by-point manner; if we may, we have numbered each point of your comments, for ease of addressing. For ease of review, we have quoted your comments in *italic* type and outlined our responses in a standard font.

Reviewer #2 (Remarks to the Author):

Dear authors

1. Thank you for the revision. I appreciate the effort you have made to make this paper clearer to read. Unfortunately, the unfamiliar HDFE (at least for me), is still something that would require clearer explanation.

I do not understand the benefit of HDFE over common linear regression. I could easily add all the covariates to a linear model and model the same association. While there is a lot of them, the sample size is huge so i think the linear model could handle them, no?

Response: Thank you for your helpful comments and for highlighting the need to clarify the benefit of high-dimensional fixed effects (HDFE) over common linear regression (such as ordinary least squares (OLS) models) for this analysis. While HDFE and OLS can produce identical estimates in simpler cases, HDFE has particular advantages when handling large numbers of fixed effects or multi-dimensionally clustered standard errors. In those cases, HDFE can adjust degrees of freedom more efficiently, providing more precise estimates (i.e., with smaller standard errors) under complex fixed-effect structures when standard errors are clustered in multi-dimensions. Additionally, HDFE offers superior computational efficiency: in our analysis (estimating the impact of individual food intake on telomere length), OLS was limited by memory requirements (Figure R1a), while HDFE completed the estimation in 45.53 seconds without any computational issues (Figure R1b). This computational efficiency allows for faster, more resource-efficient analysis, which is especially important when dealing with very large datasets and numerous covariates (—note that this efficiency gain is also important for replication purposes). Later, in addressing your fourth point below, we re-ran all models with OLS (in a more advanced computer with larger memory) and reported key estimates from both

HDFE and OLS to further illustrate their subtle differences. In the revised manuscript (page 7), we have added a few sentences to explain the modeling advantages of the HDFE model to better inform the reader.

Figure R1. Regression screenshots in Stata

a. OLS:

```
. local aaa "cookedvegetableintakeinstance0 saladrawvegetableintakeinstance0 freshfruitintakeinstance0 driedfruitin
> takeinstance0 breadintakeinstance0 cerealintakeinstance0 oilyfishintakeinstance0 nonoilyfishintakeinstance0 proce
> ssedmeatintakeinstance0 poultryintakeinstance0 beefintakeinstance0 lambmuttonintakeinstance0 porkintakeinstance0
> cheeseintakeinstance0 milkintake hotdrink coffeeintakeinstance0 teaintakeinstance0"

. reg zadjustedtsloginstance0 `aaa' male college age bmi Townsend_deprivation_index pca1-pca10 i.occupation_cat i.e
> thnic_cat i.region_cat ethnic_cat##region_cat, vce(cluster local_authority_code)
matsize too small
    You have attempted to create a matrix with too many rows or columns or attempted to fit a model with too many
    variables.  You need to increase matsize; it is currently 400.  Use set matsize; see help matsize.
```

b. HDFE:

```
. local aaa "cookedvegetableintakeinstance0 saladrawvegetableintakeinstance0 freshfruitintakeinstance0 driedfruitin
> takeinstance0 breadintakeinstance0 cerealintakeinstance0 oilyfishintakeinstance0 nonoilyfishintakeinstance0 proce
> ssedmeatintakeinstance0 poultryintakeinstance0 beefintakeinstance0 lambmuttonintakeinstance0 porkintakeinstance0
> cheeseintakeinstance0 milkintake hotdrink coffeeintakeinstance0 teaintakeinstance0"

. reghdfe zadjustedtsloginstance0 `aaa' male college age bmi Townsend_deprivation_index pca1-pca10, absorb(occupati
> on ethnic_cat region_cat ethnic_cat#region_cat) vce(cluster local_authority_code)
(dropped 13 singleton observations)
(MWFE estimator converged in 6 iterations)

HDFE Linear regression                          Number of obs   =    426,791
Absorbing 4 HDFE groups                         F(  33,    354) =     796.44
Statistics robust to heteroskedasticity         Prob > F        =     0.0000
                                                R-squared       =     0.0554
                                                Adj R-squared   =     0.0540
                                                Within R-sq.    =     0.0377
Number of clusters (local_authority_code) =       355Root MSE   =     0.9708
```

2. Are you using multiple outcomes simultaneously? That would be a step up. But this could also be done in SEM framework.

Response: Thank you for this insightful point. In our analysis, we examine the effects of multiple macronutrients—specifically carbohydrates, protein, and fat—on individual aging outcomes. While we have assessed each macronutrient's impact one at a time, we recognize that these nutrients can simultaneously influence aging, necessitating consideration of potential endogeneity and efficiency issues among them.

As you suggested, we employed a Simultaneous Equation Model (SEM) framework to enhance the robustness of our findings (Zellner, 1992; Christ et al., 2014). As detailed in Table R1, each SEM model consists of four equations: (a) the analyzed aging outcome, (b) daily carbohydrate intake, (c) daily protein intake, and (d) daily fat intake.

As reported in Table R1, the SEM estimates also indicate a potential protective effect of carbohydrate intake against aging, as evidenced by a reduction in phenotypic age (column 2a, *beta* = -0.0252, $p = 0.026$) and an increase in grey matter volume (column 3a, *beta* = 0.2615, $p = 0.008$). These results align with the MVMR findings presented in Table 2 of the manuscript, enhancing our primary conclusions. And, as reported in the manuscript, we did not observe significant associations between protein or fat intake and aging outcomes. In the revised manuscript (page 13), we have added SEM results to better inform the reader.

Table R1. SEM estimation results of macronutrients on aging

	SEM Model 1				SEM Model 2			
	1(a)	1(b)	1(c)	1(d)	2(a)	2(b)	2(c)	2(d)
	Telomere Length	Daily intake of carbohydrate	Daily intake of protein	Daily intake of fat	**Phenotypic Age**	Daily intake of carbohydrate	Daily intake of protein	Daily intake of fat
Daily intake of carbohydrate	0.0012				-0.0252**			
	(0.0021)				(0.0113)			
Daily intake of protein	0.0033				0.0087			
	(0.0094)				(0.0627)			
Daily intake of fat	-0.0060				-0.0814			
	(0.0111)				(0.0690)			
Male	-0.1733	34.3450***	8.7072***	10.9445***	5.0300***	34.0824***	8.6926***	10.9790***
	(0.1279)	(0.4183)	(0.1213)	(0.1436)	(0.6672)	(0.4476)	(0.1296)	(0.1536)
College	0.0508***	2.7059***	0.5024***	1.3936***	-0.2536***	3.0803***	0.5427***	1.4768***
	(0.0156)	(0.4282)	(0.1242)	(0.1470)	(0.0921)	(0.4585)	(0.1328)	(0.1573)
Age	-0.0229***	-0.1044***	-0.0764***	-0.1819***	0.9731***	-0.1010***	-0.0790***	-0.1796***
	(0.0016)	(0.0265)	(0.0077)	(0.0091)	(0.0090)	(0.0284)	(0.0082)	(0.0097)
BMI	-0.0054**	-0.9138***	0.3510***	0.0895***	0.2923***	-0.9079***	0.3542***	0.1037***
	(0.0026)	(0.0493)	(0.0143)	(0.0169)	(0.0159)	(0.0528)	(0.0153)	(0.0181)
Townsend deprivation index	0.0002	0.2788***	-0.2170***	0.1585***	0.0567**	0.2687***	-0.2156***	0.1543***
	(0.0037)	(0.0781)	(0.0227)	(0.0268)	(0.0236)	(0.0838)	(0.0243)	(0.0288)
rs10510554		0.3394	-0.3772***	-0.3036***		0.3023	-0.3970***	-0.3511***
		(0.2937)	(0.0852)	(0.1007)		(0.2701)	(0.0846)	(0.0909)
rs8097672		1.1889***	-0.1630	-0.0671		1.0061***	-0.2096*	-0.1805
		(0.4154)	(0.1205)	(0.1425)		(0.3784)	(0.1189)	(0.1272)
rs10206338		-0.9138***	-0.0325	0.0741		-0.8037***	-0.0146	0.0756
		(0.2945)	(0.0854)	(0.1010)		(0.2735)	(0.0851)	(0.0922)
rs10962121		0.5097*	-0.3273***	-0.1941*		0.5602**	-0.2702***	-0.1318
		(0.2914)	(0.0846)	(0.0999)		(0.2607)	(0.0829)	(0.0873)

rs2472297		-0.8065**	0.2399**	0.2821**		-0.3531	0.3853***	0.4497***
		(0.3355)	(0.0973)	(0.1150)		(0.2991)	(0.0953)	(0.1002)
rs33988101		0.8136***	-0.4611***	-0.4579***		0.8426***	-0.4248***	-0.4037***
		(0.2944)	(0.0854)	(0.1010)		(0.2822)	(0.0863)	(0.0956)
rs57193069		0.3996	0.0295	-0.1398		0.5054*	0.0857	-0.0788
		(0.2917)	(0.0846)	(0.1001)		(0.2789)	(0.0855)	(0.0944)
rs1461729		-1.7841***	0.1995	-0.0564		-2.6559***	0.0966	-0.1818
		(0.4820)	(0.1398)	(0.1654)		(0.4689)	(0.1423)	(0.1591)
rs445551		-0.2728	0.3052***	0.0967		-0.2724	0.2615***	0.0701
		(0.3140)	(0.0911)	(0.1077)		(0.2881)	(0.0902)	(0.0969)
Constant	1.3059	280.7490***	82.6645***	83.2407***	-33.1705***	280.7097***	80.6447***	81.9630***
	(1.0227)	(8.0688)	(2.3398)	(2.7704)	(5.3693)	(8.6528)	(2.5092)	(2.9682)
Top 10 PCs	Y	Y	Y	Y	Y	Y	Y	Y
Occupational FE	Y	Y	Y	Y	Y	Y	Y	Y
Ethnicity FE	Y	Y	Y	Y	Y	Y	Y	Y
Regional FE	Y	Y	Y	Y	Y	Y	Y	Y
Observations	180,142	180,142	180,142	180,142	157,020	157,020	157,020	157,020
R-squared	0.0311	0.0397	0.0376	0.0383	0.6343	0.0394	0.0379	0.0389

(Wide table to be continued)

(Wide table continued)

	SEM Model 3				SEM Model 4			
	3(a)	3(b)	3(c)	3(d)	4(a)	4(b)	4(c)	4(d)
	Grey Matter Volume	Daily intake of carbohydrate	Daily intake of protein	Daily intake of fat	**White Matter Volume**	Daily intake of carbohydrate	Daily intake of protein	Daily intake of fat
Daily intake of carbohydrate	0.2615***				-0.0440			
	(0.0991)				(0.0978)			
Daily intake of protein	0.4464				-0.7419			
	(0.8813)				(0.8699)			
Daily intake of fat	-0.3733				0.4606			
	(0.6705)				(0.6619)			
Male	-33.5630***	34.9593***	8.4549***	10.9761***	11.8103**	34.9812***	8.4604***	10.9787***
	(5.2921)	(1.0664)	(0.3055)	(0.3684)	(5.2238)	(1.0665)	(0.3055)	(0.3684)
College	-3.8523***	5.2871***	0.2385	1.7619***	-1.9687	5.2856***	0.2379	1.7619***
	(1.2719)	(1.0683)	(0.3060)	(0.3691)	(1.2555)	(1.0683)	(0.3060)	(0.3691)
Age	-3.6680***	-0.2188***	-0.0995***	-0.1818***	-1.6471***	-0.2187***	-0.0994***	-0.1818***
	(0.0822)	(0.0707)	(0.0203)	(0.0244)	(0.0811)	(0.0707)	(0.0203)	(0.0244)
BMI	-0.8455***	-1.0425***	0.3764***	0.1748***	0.2765	-1.0452***	0.3757***	0.1745***
	(0.2941)	(0.1342)	(0.0385)	(0.0464)	(0.2903)	(0.1343)	(0.0385)	(0.0464)
Townsend deprivation index	-0.2461	0.3053	-0.2079***	0.2238***	-0.3665	0.3056	-0.2078***	0.2239***
	(0.3381)	(0.2073)	(0.0594)	(0.0716)	(0.3338)	(0.2073)	(0.0594)	(0.0716)
rs10510554		-0.0435	-0.2893	-0.4520*		-0.3919	-0.3534*	-0.5056**
		(0.6814)	(0.2031)	(0.2542)		(0.7398)	(0.2078)	(0.2575)
rs8097672		-1.8321*	-0.6786**	-0.7253**		-1.3460	-0.5203*	-0.6869*
		(0.9736)	(0.2890)	(0.3602)		(1.0478)	(0.2949)	(0.3644)
rs10206338		-0.3858	-0.0148	0.2779		-0.1117	0.0722	0.3007

		(0.6801)	(0.2035)	(0.2556)		(0.7441)	(0.2087)	(0.2591)
rs10962121		1.4372**	-0.0112	-0.0373		0.8210	-0.1790	-0.1033
		(0.6593)	(0.1990)	(0.2518)		(0.7338)	(0.2050)	(0.2559)
rs2472297		-3.5381***	-0.1300	-0.0085		-3.6665***	-0.2000	-0.0038
		(0.8141)	(0.2375)	(0.2915)		(0.8461)	(0.2401)	(0.2933)
rs33988101		1.2669*	-0.2980	-0.3235		1.6372**	-0.2432	-0.2595
		(0.6848)	(0.2036)	(0.2543)		(0.7399)	(0.2081)	(0.2574)
rs57193069		-0.6841	-0.3560*	-0.4389*		-0.6524	-0.3605*	-0.4286*
		(0.6725)	(0.2012)	(0.2527)		(0.7357)	(0.2063)	(0.2562)
rs1461729		-1.4157	0.0406	-0.4258		-1.0219	0.1042	-0.3606
		(1.1341)	(0.3370)	(0.4205)		(1.2233)	(0.3442)	(0.4255)
rs445551		-0.1461	0.2090	-0.1635		-0.2283	0.2092	-0.1842
		(0.7481)	(0.2211)	(0.2746)		(0.7982)	(0.2251)	(0.2774)
Constant	925.2957***	285.9332***	100.4503***	97.8231***	827.4214***	285.3642***	100.2707***	97.7753***
	(55.5141)	(27.9885)	(8.0197)	(9.6724)	(54.7981)	(27.9987)	(8.0205)	(9.6730)
Top 10 PCs	Y	Y	Y	Y	Y	Y	Y	Y
Occupational FE	Y	Y	Y	Y	Y	Y	Y	Y
Ethnicity FE	Y	Y	Y	Y	Y	Y	Y	Y
Regional FE	Y	Y	Y	Y	Y	Y	Y	Y
Observations	24,875	24,875	24,875	24,875	24,875	24,875	24,875	24,875
R-squared	0.2366	0.0499	0.0423	0.0446	-0.0285	0.0499	0.0423	0.0447

Note: Standard errors are in parentheses. Asterisks represent statistical significance at *** p<0.01, ** p<0.05, and * p<0.1, respectively.

References:

[1] Zellner, A., & Theil, H. (1992). Three-stage least squares: simultaneous estimation of simultaneous equations. In *Henri Theil's contributions to economics and econometrics: econometric theory and methodology* (pp. 147-178). Dordrecht: Springer Netherlands.

[2] Christ, S. L., Lee, D. J., Lam, B. L., & Zheng, D. D. (2014). Structural equation modeling: a framework for ocular and other medical sciences research. *Ophthalmic epidemiology*, *21*(1), 1-13.

3. Are the statistical effects independent of all other predictors & covariates?

Response: Yes, the statistical effects have been adjusted for all other predictors and covariates in our analysis. This adjustment ensures that the observed relationships are independent of these additional factors, enhancing the robustness of our findings.

4. I am not statistically versed nor do I have the time to read the HDFE paper in detail to understand how it differs from regular linear model. Since you bring a new method to the table, you have to convince us, non-statisticians that it brings something more to the table.

Response: Thank you for expressing your concerns about the suitability and added value of the HDFE model in our study. We understand that as this approach may be unfamiliar to some readers, it is essential to clearly demonstrate its benefits over conventional linear regression models (e.g., OLS). As noted in our response to your first point above, the HDFE model offers distinct advantages, particularly when dealing with a large number of fixed effects and multi-dimensionally clustered standard errors.

Unlike OLS, which can become computationally inefficient and less precise (i.e., with larger standard errors) when accounting for hundreds or thousands of fixed effects, HDFE is specifically designed to handle high-dimensional fixed effects efficiently. It adjusts for the degrees of freedom in a manner that ensures unbiased estimation of standard errors, even in cases where there is a need to cluster standard errors across multiple dimensions. Additionally, HDFE is computationally faster and less memory-intensive than OLS with dummy variables, making it particularly well-suited for analyzing large datasets like ours.

To further illustrate these points and address your concerns, we have included a direct comparison of the results obtained using HDFE and OLS in response to your next comment (Table R2). The comparison shows that while the coefficient estimates are nearly identical, the HDFE model consistently delivers more precise estimates (with smaller p-values) due to its superior estimation efficiency. We hope this explanation and

the accompanying results convincingly demonstrate the value of using HDFE in our analysis.

5. I would also like to see the same associations modelled with a traditional linear regression to see, whether there is in fact any differnece between HDFE and linear regression.

Response: Thank you for your suggestion. We have re-estimated the model using a traditional linear regression approach (OLS with dummy variables) for comparison. In Table R2, we present both HDFE and OLS estimates (beta and *p*-values) side by side for each of the four aging outcomes. The results indicate that the estimated coefficients (beta) from HDFE and OLS are nearly identical to the fourth decimal place. However, the raw *p*-values from OLS are generally larger than those from HDFE, reflecting reduced estimation efficiency in OLS, as noted in our response to your earlier comments. This comparison underscores the advantage of HDFE in terms of computational efficiency and statistical precision.

Table R2. HDFE and OLS estimation results of dietary factors on aging

(1) Outcome: Telomere Length

	HDFE Model 1		OLS Model 1	
	Outcome Variable: Telomere Length		Outcome Variable: Telomere Length	
Exposure	*Beta*	*p value*	*Beta*	*p value*
(A) Food intake				
Cooked vegetable	0.0015	0.0933	0.0015	0.1068
Salad/raw vegetable	-0.0008	0.2730	-0.0008	0.3369
Fresh fruit	0.0058	< 0.0001	0.0058	< 0.0001
Dried fruit	0.0050	< 0.0001	0.0050	< 0.0001
Bread	0.0003	0.2608	0.0003	0.3443
Cereal	0.0035	< 0.0001	0.0035	< 0.0001
Oily fish	0.0073	< 0.0001	0.0073	< 0.0001
Non-oily fish	-0.0009	0.6194	-0.0009	0.6046
Processed meat	-0.0044	0.0001	-0.0044	0.0003
Poultry	-0.0037	0.0027	-0.0037	0.0035
Beef	-0.0039	0.0303	-0.0039	0.0929
Lamb/mutton	0.0030	0.2467	0.0030	0.2885
Pork	-0.0005	0.7839	-0.0005	0.8627
Cheese	0.0048	< 0.0001	0.0048	< 0.0001
Milk	-0.0034	0.5091	-0.0034	0.5733
Hot drink	-0.0034	0.4070	-0.0034	0.4452
Coffee	-0.0041	< 0.0001	-0.0041	< 0.0001
Tea	-0.0020	0.0003	-0.0020	0.0006
(B) Dietary patterns				
Low carb diet	-0.0191	0.0002	-0.0191	0.0004
High carb diet	0.0117	0.9599	0.0117	0.8966
High protein diet	-0.0092	0.0919	-0.0092	0.1993
Low fat diet	-0.0131	0.0104	-0.0131	0.0248
High fat diet	-0.0019	0.0400	-0.0019	0.9365
Ketogenic diet	-0.0918	0.1130	-0.0918	0.1515
(C) Macronutrients				
Daily intake of carbohydrate	0.0001	0.0058	0.0001	0.0076
Daily intake of protein	-0.0001	0.4711	-0.0001	0.5254
Daily intake of fat	-0.0000	0.6300	-0.0000	0.6255
(D) Diet quality scores				
HEI-2015 score	0.0012	< 0.0001	0.0012	< 0.0001
DASH score	0.0021	0.2399	0.0021	0.3055
Food diversity index	0.0077	< 0.0001	0.0078	< 0.0001

(2) Outcome: Phenotypic Age

	HDFE Model 2		OLS Model 2	
	Outcome Variable: Phenotypic Age		Outcome Variable: Phenotypic Age	
Exposure	*Beta*	*p value*	*Beta*	*p value*
(A) Food intake				
Cooked vegetable	-0.0314	< 0.0001	-0.0314	< 0.0001
Salad/raw vegetable	-0.0377	< 0.0001	-0.0377	< 0.0001
Fresh fruit	-0.0968	< 0.0001	-0.0968	< 0.0001
Dried fruit	-0.0602	< 0.0001	-0.0602	< 0.0001
Bread	0.0077	< 0.0001	0.0077	< 0.0001
Cereal	-0.0567	< 0.0001	-0.0567	< 0.0001
Oily fish	-0.0971	< 0.0001	-0.0971	< 0.0001
Non-oily fish	-0.0240	0.0025	-0.0240	0.0039
Processed meat	0.1389	< 0.0001	0.1389	< 0.0001
Poultry	0.0174	0.0045	0.0174	0.0065
Beef	0.0790	< 0.0001	0.0790	< 0.0001
Lamb/mutton	0.1097	< 0.0001	0.1097	< 0.0001
Pork	0.1569	< 0.0001	0.1569	< 0.0001
Cheese	-0.0585	< 0.0001	-0.0585	< 0.0001
Milk	0.2353	< 0.0001	0.2353	< 0.0001
Hot drink	-0.3214	< 0.0001	-0.3214	< 0.0001
Coffee	-0.0469	< 0.0001	-0.0469	< 0.0001
Tea	0.0128	< 0.0001	0.0128	< 0.0001
(B) Dietary patterns				
Low carb diet	0.0615	0.0082	0.0615	0.0170
High carb diet	0.0176	0.5962	0.0176	0.8010
High protein diet	0.0969	0.0072	0.0969	0.0013
Low fat diet	-0.2236	< 0.0001	-0.2236	< 0.0001
High fat diet	0.1857	0.5458	0.1857	0.5988
Ketogenic diet	-0.1896	0.7559	-0.1896	0.8557
(C) Macronutrients				
Daily intake of carbohydrate	-0.0012	< 0.0001	-0.0012	< 0.0001
Daily intake of protein	-0.0007	0.1491	-0.0007	0.1578
Daily intake of fat	0.0042	< 0.0001	0.0042	< 0.0001
(D) Diet quality scores				
HEI-2015 score	-0.0274	< 0.0001	-0.0274	< 0.0001
DASH score	-0.1229	< 0.0001	-0.1229	< 0.0001
Food diversity index	-0.1158	< 0.0001	-0.1158	< 0.0001

(3) Outcome: Grey Matter Volume

	HDFE Model 3		OLS Model 3	
	Outcome Variable: Grey Matter Volume		Outcome Variable: Grey Matter Volume	
Exposure	*Beta*	*p value*	*Beta*	*p value*
(A) Food intake				
Cooked vegetable	-0.1255	0.2339	-0.1255	0.2565
Salad/raw vegetable	-0.0826	0.4351	-0.0826	0.5116
Fresh fruit	-0.3759	0.0107	-0.3759	0.0273
Dried fruit	-0.0612	0.4543	-0.0612	0.6198
Bread	0.0687	0.0091	0.0687	0.0053
Cereal	0.5701	< 0.0001	0.5701	< 0.0001
Oily fish	-1.1153	< 0.0001	-1.1153	< 0.0001
Non-oily fish	-0.2739	0.1152	-0.2739	0.1344
Processed meat	-0.3106	0.0470	-0.3106	0.0429
Poultry	0.6144	0.0001	0.6144	0.0002
Beef	-0.0620	0.6298	-0.0620	0.7768
Lamb/mutton	-1.3125	0.0015	-1.3125	0.0032
Pork	-0.2674	0.3937	-0.2674	0.4524
Cheese	-0.6096	< 0.0001	-0.6096	< 0.0001
Milk	-0.0342	0.9095	-0.0342	0.9529
Hot drink	-0.3675	0.5667	-0.3675	0.7932
Coffee	-0.9431	< 0.0001	-0.9431	< 0.0001
Tea	-0.0650	0.4665	-0.0650	0.3562
(B) Dietary patterns				
Low carb diet	-5.5158	< 0.0001	-5.5158	< 0.0001
High carb diet	3.6401	0.5402	3.6401	0.6289
High protein diet	2.0250	0.0192	2.0250	0.0216
Low fat diet	-2.5276	< 0.0001	-2.5276	< 0.0001
High fat diet	0.3949	0.0001	0.3949	0.8886
Ketogenic diet	-14.8264	0.0376	-14.8264	0.0572
(C) Macronutrients				
Daily intake of carbohydrate	0.0198	< 0.0001	0.0198	< 0.0001
Daily intake of protein	-0.0269	0.0293	-0.0269	0.0316
Daily intake of fat	-0.0270	0.0111	-0.0265	0.0193
(D) Diet quality scores				
HEI-2015 score	-0.0042	0.8369	-0.0042	0.8720
DASH score	-0.0915	0.6232	-0.0915	0.6239
Food diversity index	-0.1681	0.2478	-0.1681	0.3396

(4) Outcome: White Matter Volume

	HDFE Model 4		OLS Model 4	
	Outcome Variable: White Matter Volume		Outcome Variable: White Matter Volume	
Exposure	*Beta*	*p value*	*Beta*	*p value*
(A) Food intake				
Cooked vegetable	0.0316	0.6578	0.0316	0.7717
Salad/raw vegetable	0.1407	0.1975	0.1407	0.2097
Fresh fruit	0.1679	0.2005	0.1679	0.2662
Dried fruit	-0.1132	0.2848	-0.1132	0.3101
Bread	0.0612	0.0162	0.0612	0.0142
Cereal	0.2383	0.0023	0.2383	0.0025
Oily fish	-0.3675	0.0803	-0.3675	0.0952
Non-oily fish	-0.1176	0.7466	-0.1176	0.8245
Processed meat	0.0999	0.6166	0.0999	0.7018
Poultry	0.3141	0.0572	0.3141	0.0922
Beef	-0.2494	0.2960	-0.2494	0.3185
Lamb/mutton	-0.8336	0.0981	-0.8336	0.0502
Pork	-0.2993	0.3582	-0.2993	0.4562
Cheese	-0.3637	0.0006	-0.3637	0.0011
Milk	0.4389	0.5495	0.4389	0.5381
Hot drink	0.9637	0.0456	0.9637	0.1236
Coffee	-0.2523	0.0357	-0.2523	0.0314
Tea	0.1687	0.0190	0.1687	0.0361
(B) Dietary patterns				
Low carb diet	-2.3143	< 0.0001	-2.3143	< 0.0001
High carb diet	1.2110	0.2500	1.2110	0.5285
High protein diet	2.4727	0.0005	2.4727	0.0014
Low fat diet	-0.7493	0.1743	-0.7493	0.1442
High fat diet	-5.7327	0.0017	-5.7327	0.1001
Ketogenic diet	-6.5673	0.4022	-6.5673	0.5013
(C) Macronutrients				
Daily intake of carbohydrate	0.0100	0.0057	0.0100	0.0138
Daily intake of protein	0.0095	0.4756	0.0095	0.4942
Daily intake of fat	-0.0297	0.0088	-0.0297	0.0056
(D) Diet quality scores				
HEI-2015 score	-0.0057	0.7946	-0.0057	0.8410
DASH score	0.1012	0.6159	0.1012	0.5814
Food diversity index	-0.3120	0.0479	-0.3120	0.0619

6. I think the 477 FEs covariates should also be listed. The best would be to share analysis code, so people wishing to do so, could reproduce the code.

Response: Thank you for your valuable suggestion. We have taken your feedback into account and included a detailed list of the 477 fixed effects in Table R3 at the end of this response letter, including 336 occupational FEs, 18 race/ethnicity FEs, 10 regional FEs, and 113 race/ethnicity-by-region FEs. Additionally, to facilitate reproducibility, we have uploaded the main code for running the HDFE and OLS regressions, including all fixed effects, to the following GitHub repository: https://github.com/maxwell2732/diet_and_aging/. The specific file is titled Diet_Aging_HDFE_OLS_wFE.do. We hope this provides clarity and enables further exploration and replication of our analysis.

7. I understand that some of the FEs were also used in MR. If you list them, then clarify, which are in MR and which not.

Response: This is an excellent point. Mendelian Randomization (MR) is a causal inference approach that leverages genetic variants as instrumental variables, thereby mitigating confounding bias. As such, it is not necessary to include as many fixed effects in MVMR analyses as in HDFE or OLS regressions. For our MVMR analyses, we included a more focused set of 18 race/ethnicity fixed effects and 10 regional fixed effects to account for key demographic and geographic variations. A detailed list of these 28 fixed effects has been provided in Table R4 at the end of this response letter. We hope this clarification addresses your concerns.

Table R3. List of fixed effects included in HDFE and OLS regressions

No.	Fixed effects	Description
1	2.occupation_cat	Carpenters and joiners
2	3.occupation_cat	Office managers
3	4.occupation_cat	IT operations technicians
4	5.occupation_cat	Marketing and sales managers
5	6.occupation_cat	Secondary education teaching professionals
6	7.occupation_cat	Psychologists
7	8.occupation_cat	Software professionals
8	9.occupation_cat	Security guards and related occupations
9	10.occupation_cat	Medical secretaries
10	11.occupation_cat	Market research interviewers
11	12.occupation_cat	Higher education teaching professionals
12	13.occupation_cat	Sheet metal workers
13	14.occupation_cat	Garage managers and proprietors
14	15.occupation_cat	Advertising and public relations managers
15	16.occupation_cat	Prison service officers (below principal officer)
16	17.occupation_cat	Nurses
17	18.occupation_cat	Public service associate professionals
18	19.occupation_cat	Local government clerical officers and assistants
19	20.occupation_cat	Cleaners, domestics
20	21.occupation_cat	Personal assistants and other secretaries
21	22.occupation_cat	Production, works and maintenance managers
22	23.occupation_cat	Financial managers and chartered secretaries
23	24.occupation_cat	Legal secretaries
24	25.occupation_cat	Taxation experts
25	26.occupation_cat	Butchers, meat cutters
26	27.occupation_cat	Medical practitioners
27	28.occupation_cat	Printing machine minders and assistants
28	29.occupation_cat	General office assistants/clerks
29	30.occupation_cat	Civil Service administrative officers and assistants
30	31.occupation_cat	Metal working production and maintenance fitters
31	32.occupation_cat	Educational assistants
32	33.occupation_cat	Care assistants and home carers
33	34.occupation_cat	Personnel and industrial relations officers
34	35.occupation_cat	Accounts and wages clerks, book-keepers, other financial clerks
35	36.occupation_cat	Nursing auxiliaries and assistants
36	37.occupation_cat	Sales and retail assistants
37	38.occupation_cat	Graphic designers
38	39.occupation_cat	Planning and quality control engineers
39	40.occupation_cat	Primary and nursery education teaching professionals
40	41.occupation_cat	Teaching professionals n.e.c.
41	42.occupation_cat	Product, clothing and related designers
42	43.occupation_cat	Tool makers, tool fitters and markers-out
43	44.occupation_cat	Business and related associate professionals n.e.c.
44	45.occupation_cat	Housing and welfare officers
45	46.occupation_cat	Healthcare practice managers
46	47.occupation_cat	Storage and warehouse managers
47	48.occupation_cat	Authors, writers

48	49.occupation_cat	Personnel, training and industrial relations managers
49	50.occupation_cat	Environmental health officers
50	51.occupation_cat	Other goods handling and storage occupations n.e.c.
51	52.occupation_cat	Transport and distribution managers
52	53.occupation_cat	Civil Service executive officers
53	54.occupation_cat	Quality assurance managers
54	55.occupation_cat	Kitchen and catering assistants
55	56.occupation_cat	Clergy
56	57.occupation_cat	Van drivers
57	58.occupation_cat	Biological scientists and biochemists
58	59.occupation_cat	Solicitors and lawyers, judges and coroners
59	60.occupation_cat	Chemical engineers
60	61.occupation_cat	Air travel assistants
61	62.occupation_cat	Farmers
62	63.occupation_cat	Quality assurance technicians
63	64.occupation_cat	Information and communication technology managers
64	65.occupation_cat	Nursery nurses
65	66.occupation_cat	Chartered and certified accountants
66	67.occupation_cat	Filing and other records assistants/clerks
67	68.occupation_cat	Childminders and related occupations
68	69.occupation_cat	Broadcasting associate professionals
69	70.occupation_cat	Construction trades n.e.c.
70	71.occupation_cat	Heavy goods vehicle drivers
71	72.occupation_cat	Electrical/electronics engineers n.e.c.
72	73.occupation_cat	Travel agents
73	74.occupation_cat	Hospital and health service managers
74	75.occupation_cat	Public service administrative professionals
75	76.occupation_cat	Credit controllers
76	77.occupation_cat	Buyers and purchasing officers
77	78.occupation_cat	Social services managers
78	79.occupation_cat	Receptionists
79	80.occupation_cat	Sales representatives
80	81.occupation_cat	Restaurant and catering managers
81	82.occupation_cat	Chefs, cooks
82	83.occupation_cat	Senior officials in national government
83	84.occupation_cat	Midwives
84	85.occupation_cat	Hairdressers, barbers
85	86.occupation_cat	Farm workers
86	87.occupation_cat	Metal machining setters and setter-operators
87	88.occupation_cat	Design and development engineers
88	89.occupation_cat	IT strategy and planning professionals
89	90.occupation_cat	Police officers (sergeant and below)
90	91.occupation_cat	Labourers in process and plant operations n.e.c.
91	92.occupation_cat	Legal associate professionals
92	93.occupation_cat	Management consultants, actuaries, economists and statisticians
93	94.occupation_cat	Careers advisers and vocational guidance specialists
94	95.occupation_cat	Civil engineers
95	96.occupation_cat	Finance and investment analysts/advisers
96	97.occupation_cat	Housekeepers and related occupations
97	98.occupation_cat	Gardeners and groundsmen/groundswomen

98	99.occupation_cat	Registrars and senior administrators of educational establishments
99	100.occupation_cat	Senior officials of special interest organisations
100	101.occupation_cat	Electricians, electrical fitters
101	102.occupation_cat	Publicans and managers of licensed premises
102	103.occupation_cat	Journalists, newspaper and periodical editors
103	104.occupation_cat	Air transport operatives
104	105.occupation_cat	Driving instructors
105	106.occupation_cat	Customer care occupations
106	107.occupation_cat	Photographers and audio-visual equipment operators
107	108.occupation_cat	Scientific researchers
108	109.occupation_cat	Rail transport operatives
109	110.occupation_cat	Town planners
110	111.occupation_cat	Postal workers, mail sorters, messengers, couriers
111	112.occupation_cat	Managers in construction
112	113.occupation_cat	Debt, rent and other cash collectors
113	114.occupation_cat	Leisure and theme park attendants
114	115.occupation_cat	Taxi, cab drivers and chauffeurs
115	116.occupation_cat	Financial institution managers
116	117.occupation_cat	Mechanical engineers
117	118.occupation_cat	Leisure and sports managers
118	119.occupation_cat	Houseparents and residential wardens
119	120.occupation_cat	Metal making and treating process operatives
120	121.occupation_cat	Chartered surveyors (not quantity surveyors)
121	122.occupation_cat	Property, housing and land managers
122	123.occupation_cat	Senior officials in local government
123	124.occupation_cat	Waiters, waitresses
124	125.occupation_cat	Plastics process operatives
125	126.occupation_cat	Managers and proprietors in other services n.e.c.
126	127.occupation_cat	Social workers
127	128.occupation_cat	Plumbers, heating and ventilating engineers
128	129.occupation_cat	Quantity surveyors
129	130.occupation_cat	Glaziers, window fabricators and fitters
130	131.occupation_cat	Management accountants
131	132.occupation_cat	Customer care managers
132	133.occupation_cat	Library assistants/clerks
133	134.occupation_cat	Dental nurses
134	135.occupation_cat	Architectural technologists and town planning technicians
135	136.occupation_cat	Lines repairers and cable jointers
136	137.occupation_cat	Computer engineers, installation and maintenance
137	138.occupation_cat	Chiropodists
138	139.occupation_cat	School secretaries
139	140.occupation_cat	Therapists n.e.c.
140	141.occupation_cat	Leisure and travel service occupations n.e.c.
141	142.occupation_cat	Metal working machine operatives
142	143.occupation_cat	Company secretaries
143	144.occupation_cat	Road construction operatives
144	145.occupation_cat	Electrical/electronics technicians
145	146.occupation_cat	Painters and decorators
146	147.occupation_cat	Caretakers
147	148.occupation_cat	Welding trades

148	149.occupation_cat	Statutory examiners
149	150.occupation_cat	Researchers n.e.c.
150	151.occupation_cat	Draughtspersons
151	152.occupation_cat	Further education teaching professionals
152	153.occupation_cat	Librarians
153	154.occupation_cat	Elementary sales occupations n.e.c.
154	155.occupation_cat	Conference and exhibition managers
155	156.occupation_cat	Traffic wardens
156	157.occupation_cat	Physiotherapists
157	158.occupation_cat	Retail and wholesale managers
158	159.occupation_cat	Special needs education teaching professionals
159	160.occupation_cat	Pharmacists/pharmacologists
160	161.occupation_cat	Inspectors of factories, utilities and trading standards
161	162.occupation_cat	Hand craft occupations n.e.c.
162	163.occupation_cat	Car park attendants
163	164.occupation_cat	Directors and chief executives of major organisations
164	165.occupation_cat	Food, drink and tobacco process operatives
165	166.occupation_cat	Residential and day care managers
166	167.occupation_cat	Plasterers
167	168.occupation_cat	Financial and accounting technicians
168	169.occupation_cat	Fire service officers (leading fire officer and below)
169	170.occupation_cat	Precision instrument makers and repairers
170	171.occupation_cat	Train drivers
171	172.occupation_cat	Veterinarians
172	173.occupation_cat	Artists
173	174.occupation_cat	Estate agents, auctioneers
174	175.occupation_cat	Architects
175	176.occupation_cat	Seafarers (merchant navy); barge, lighter and boat operatives
176	177.occupation_cat	Arts officers, producers and directors
177	178.occupation_cat	Bar staff
178	179.occupation_cat	Communication operators
179	180.occupation_cat	Production and process engineers
180	181.occupation_cat	Chemical and related process operatives
181	182.occupation_cat	Sports coaches, instructors and officials
182	183.occupation_cat	Public relations officers
183	184.occupation_cat	Call centre agents/operators
184	185.occupation_cat	Laboratory technicians
185	186.occupation_cat	Chemists
186	187.occupation_cat	Engineering professionals n.e.c.
187	188.occupation_cat	School mid-day assistants
188	189.occupation_cat	Education officers, school inspectors
189	190.occupation_cat	Medical and dental technicians
190	191.occupation_cat	Countryside and park rangers
191	192.occupation_cat	Labourers in building and woodworking trades
192	193.occupation_cat	Retail cashiers and check-out operators
193	194.occupation_cat	Stevedores, dockers and slingers
194	195.occupation_cat	Marketing associate professionals
195	196.occupation_cat	Officers of non-governmental organisations
196	197.occupation_cat	Mobile machine drivers and operatives n.e.c.
197	198.occupation_cat	Hairdressing and beauty salon managers and proprietors

198	199.occupation_cat	Packers, bottlers, canners, fillers
199	200.occupation_cat	Estimators, valuers and assessors
200	201.occupation_cat	Brokers
201	202.occupation_cat	Vocational and industrial trainers and instructors
202	203.occupation_cat	Pharmaceutical dispensers
203	204.occupation_cat	Shopkeepers and wholesale/retail dealers
204	205.occupation_cat	Elementary office occupations n.e.c.
205	206.occupation_cat	Electronics engineers
206	207.occupation_cat	Bus and coach drivers
207	208.occupation_cat	Actors, entertainers
208	209.occupation_cat	Occupational hygienists and safety officers (health and safety)
209	210.occupation_cat	Hotel and accommodation managers
210	211.occupation_cat	Physicists, geologists and meteorologists
211	212.occupation_cat	Telecommunications engineers
212	213.occupation_cat	Stock control clerks
213	214.occupation_cat	Typists
214	215.occupation_cat	Sports and fitness occupations n.e.c.
215	216.occupation_cat	Sales related occupations n.e.c.
216	217.occupation_cat	Counter clerks
217	218.occupation_cat	Engineering technicians
218	219.occupation_cat	Transport and distribution clerks
219	220.occupation_cat	Speech and language therapists
220	221.occupation_cat	Pensions and insurance clerks
221	222.occupation_cat	Research and development managers
222	223.occupation_cat	Vehicle body builders and repairers
223	224.occupation_cat	Printers
224	225.occupation_cat	Medical radiographers
225	226.occupation_cat	Upholsterers
226	227.occupation_cat	Probation officers
227	228.occupation_cat	Telephone salespersons
228	229.occupation_cat	Senior officers in fire, ambulance, prison and related services
229	230.occupation_cat	Motor mechanics, auto engineers
230	231.occupation_cat	Purchasing managers
231	232.occupation_cat	Telephonists
232	233.occupation_cat	Dental practitioners
233	234.occupation_cat	Construction operatives n.e.c.
234	235.occupation_cat	Protective service associate professionals n.e.c.
235	236.occupation_cat	Forestry workers
236	237.occupation_cat	TV, video and audio engineers
237	238.occupation_cat	Occupational therapists
238	239.occupation_cat	Animal care occupations n.e.c.
239	240.occupation_cat	Musicians
240	241.occupation_cat	Beauticians and related occupations
241	242.occupation_cat	Plant and machine operatives n.e.c.
242	243.occupation_cat	Science and engineering technicians n.e.c.
243	244.occupation_cat	Roundsmen/women and van salespersons
244	245.occupation_cat	Dancers and choreographers
245	246.occupation_cat	Farm managers
246	247.occupation_cat	Metal plate workers, shipwrights, riveters
247	248.occupation_cat	Rubber process operatives

248	249.occupation_cat	Crane drivers
249	250.occupation_cat	Bricklayers, masons
250	251.occupation_cat	Ambulance staff (excluding paramedics)
251	252.occupation_cat	Security managers
252	253.occupation_cat	Police officers (inspectors and above)
253	254.occupation_cat	Playgroup leaders/assistants
254	255.occupation_cat	Merchandisers and window dressers
255	256.occupation_cat	Youth and community workers
256	257.occupation_cat	Conservation and environmental protection officers
257	258.occupation_cat	Vehicle spray painters
258	259.occupation_cat	Labourers in other construction trades n.e.c.
259	260.occupation_cat	Legal professionals n.e.c.
260	261.occupation_cat	Launderers, dry cleaners, pressers
261	262.occupation_cat	Screen printers
262	263.occupation_cat	Coal mine operatives
263	264.occupation_cat	Database assistants/clerks
264	265.occupation_cat	Insurance underwriters
265	266.occupation_cat	Archivists and curators
266	267.occupation_cat	Rail travel assistants
267	268.occupation_cat	Assemblers (vehicles and metal goods)
268	269.occupation_cat	Bakers, flour confectioners
269	270.occupation_cat	Social science researchers
270	271.occupation_cat	Routine inspectors and testers
271	272.occupation_cat	Roofers, roof tilers and slaters
272	273.occupation_cat	Electrical engineers
273	274.occupation_cat	Goldsmiths, silversmiths, precious stone workers
274	275.occupation_cat	Horticultural trades
275	276.occupation_cat	Fork-lift truck drivers
276	277.occupation_cat	Sewing machinists
277	278.occupation_cat	Building and civil engineering technicians
278	279.occupation_cat	Veterinary nurses and assistants
279	280.occupation_cat	Fitness instructors
280	281.occupation_cat	Window cleaners
281	282.occupation_cat	Routine laboratory testers
282	283.occupation_cat	Ophthalmic opticians
283	284.occupation_cat	Pipe fitters
284	285.occupation_cat	Ship and hovercraft officers
285	286.occupation_cat	Scaffolders, stagers, riggers
286	287.occupation_cat	Agricultural and fishing trades n.e.c.
287	288.occupation_cat	Fishing and agriculture related occupations n.e.c.
288	289.occupation_cat	Travel agency managers
289	290.occupation_cat	Industrial cleaning process occupations
290	291.occupation_cat	Assemblers and routine operatives n.e.c.
291	292.occupation_cat	Paper and wood machine operatives
292	293.occupation_cat	Energy plant operatives
293	294.occupation_cat	Pharmacy managers
294	295.occupation_cat	Fishmongers, poultry dressers
295	296.occupation_cat	Aircraft pilots and flight engineers
296	297.occupation_cat	Collector salespersons and credit agents
297	298.occupation_cat	Tailors and dressmakers

298	299.occupation_cat	Musical instrument makers and tuners
299	300.occupation_cat	Dispensing opticians
300	301.occupation_cat	Furniture makers, other craft woodworkers
301	302.occupation_cat	Glass and ceramics process operatives
302	303.occupation_cat	Agricultural machinery drivers
303	304.occupation_cat	IT user support technicians
304	305.occupation_cat	Pest control officers
305	306.occupation_cat	Managers in animal husbandry, forestry and fishing n.e.c.
306	307.occupation_cat	Assemblers (electrical products)
307	308.occupation_cat	Bookbinders and print finishers
308	309.occupation_cat	Floorers and wall tilers
309	310.occupation_cat	Elementary personal services occupations n.e.c.
310	311.occupation_cat	Officers in armed forces
311	312.occupation_cat	Textiles, garments and related trades n.e.c.
312	313.occupation_cat	Shelf fillers
313	314.occupation_cat	Floral arrangers, florists
314	315.occupation_cat	Importers, exporters
315	316.occupation_cat	Smiths and forge workers
316	317.occupation_cat	Hospital porters
317	318.occupation_cat	Market and street traders and assistants
318	319.occupation_cat	Undertakers and mortuary assistants
319	320.occupation_cat	Pattern makers (moulds)
320	321.occupation_cat	Rail construction and maintenance operatives
321	322.occupation_cat	Water and sewerage plant operatives
322	323.occupation_cat	Process operatives n.e.c.
323	324.occupation_cat	Recycling and refuse disposal managers
324	325.occupation_cat	Textile process operatives
325	326.occupation_cat	Elementary security occupations n.e.c.
326	327.occupation_cat	School crossing patrol attendants
327	328.occupation_cat	Transport operatives n.e.c.
328	329.occupation_cat	Road sweepers
329	330.occupation_cat	Paramedics
330	331.occupation_cat	Natural environment and conservation managers
331	332.occupation_cat	Travel and tour guides
332	333.occupation_cat	Glass and ceramics makers, decorators and finishers
333	334.occupation_cat	Sports players
334	335.occupation_cat	Weighers, graders, sorters
335	336.occupation_cat	Refuse and salvage occupations
336	337.occupation_cat	Sports and leisure assistants
337	2.ethnic_cat	African
338	3.ethnic_cat	Any other Asian background
339	4.ethnic_cat	Any other Black background
340	5.ethnic_cat	Any other mixed background
341	6.ethnic_cat	Any other white background
342	7.ethnic_cat	Bangladeshi
343	8.ethnic_cat	British
344	9.ethnic_cat	Caribbean
345	10.ethnic_cat	Chinese
346	11.ethnic_cat	Indian
347	12.ethnic_cat	Irish

348	13.ethnic_cat	Other ethnic group
349	14.ethnic_cat	Pakistani
350	15.ethnic_cat	Mixed
351	16.ethnic_cat	White
352	17.ethnic_cat	White and Asian
353	18.ethnic_cat	White and Black African
354	19.ethnic_cat	White and Black Caribbean
355	2.region_cat	East of England
356	3.region_cat	London
357	4.region_cat	North East
358	5.region_cat	North West
359	6.region_cat	Scotland
360	7.region_cat	South East
361	8.region_cat	South West
362	9.region_cat	Wales
363	10.region_cat	West Midlands
364	11.region_cat	Yorkshire and The Humber
365	2.ethnic_cat#2.region_cat	-
366	2.ethnic_cat#5.region_cat	-
367	2.ethnic_cat#7.region_cat	-
368	2.ethnic_cat#8.region_cat	-
369	2.ethnic_cat#9.region_cat	-
370	2.ethnic_cat#11.region_cat	-
371	3.ethnic_cat#3.region_cat	-
372	3.ethnic_cat#4.region_cat	-
373	3.ethnic_cat#5.region_cat	-
374	3.ethnic_cat#6.region_cat	-
375	3.ethnic_cat#8.region_cat	-
376	3.ethnic_cat#9.region_cat	-
377	3.ethnic_cat#11.region_cat	-
378	4.ethnic_cat#2.region_cat	-
379	4.ethnic_cat#3.region_cat	-
380	4.ethnic_cat#5.region_cat	-
381	4.ethnic_cat#6.region_cat	-
382	4.ethnic_cat#7.region_cat	-
383	4.ethnic_cat#8.region_cat	-
384	4.ethnic_cat#11.region_cat	-
385	5.ethnic_cat#3.region_cat	-
386	5.ethnic_cat#5.region_cat	-
387	5.ethnic_cat#6.region_cat	-
388	5.ethnic_cat#8.region_cat	-
389	5.ethnic_cat#10.region_cat	-
390	5.ethnic_cat#11.region_cat	-
391	6.ethnic_cat#3.region_cat	-
392	6.ethnic_cat#4.region_cat	-
393	6.ethnic_cat#7.region_cat	-
394	6.ethnic_cat#10.region_cat	-
395	6.ethnic_cat#11.region_cat	-
396	7.ethnic_cat#3.region_cat	-
397	7.ethnic_cat#4.region_cat	-

398	7.ethnic_cat#5.region_cat	-
399	7.ethnic_cat#8.region_cat	-
400	7.ethnic_cat#9.region_cat	-
401	7.ethnic_cat#11.region_cat	-
402	8.ethnic_cat#3.region_cat	-
403	8.ethnic_cat#7.region_cat	-
404	8.ethnic_cat#10.region_cat	-
405	8.ethnic_cat#11.region_cat	-
406	9.ethnic_cat#2.region_cat	-
407	9.ethnic_cat#3.region_cat	-
408	9.ethnic_cat#5.region_cat	-
409	9.ethnic_cat#7.region_cat	-
410	9.ethnic_cat#8.region_cat	-
411	9.ethnic_cat#10.region_cat	-
412	9.ethnic_cat#11.region_cat	-
413	10.ethnic_cat#2.region_cat	-
414	10.ethnic_cat#3.region_cat	-
415	10.ethnic_cat#6.region_cat	-
416	10.ethnic_cat#7.region_cat	-
417	10.ethnic_cat#8.region_cat	-
418	10.ethnic_cat#10.region_cat	-
419	11.ethnic_cat#2.region_cat	-
420	11.ethnic_cat#3.region_cat	-
421	11.ethnic_cat#5.region_cat	-
422	11.ethnic_cat#6.region_cat	-
423	11.ethnic_cat#8.region_cat	-
424	11.ethnic_cat#10.region_cat	-
425	11.ethnic_cat#11.region_cat	-
426	12.ethnic_cat#3.region_cat	-
427	12.ethnic_cat#4.region_cat	-
428	12.ethnic_cat#6.region_cat	-
429	12.ethnic_cat#8.region_cat	-
430	12.ethnic_cat#10.region_cat	-
431	12.ethnic_cat#11.region_cat	-
432	13.ethnic_cat#2.region_cat	-
433	13.ethnic_cat#3.region_cat	-
434	13.ethnic_cat#6.region_cat	-
435	13.ethnic_cat#7.region_cat	-
436	13.ethnic_cat#9.region_cat	-
437	13.ethnic_cat#11.region_cat	-
438	14.ethnic_cat#2.region_cat	-
439	14.ethnic_cat#3.region_cat	-
440	14.ethnic_cat#4.region_cat	-
441	14.ethnic_cat#7.region_cat	-
442	14.ethnic_cat#9.region_cat	-
443	14.ethnic_cat#10.region_cat	-
444	15.ethnic_cat#2.region_cat	-
445	15.ethnic_cat#3.region_cat	-
446	15.ethnic_cat#5.region_cat	-
447	15.ethnic_cat#7.region_cat	-

448	15.ethnic_cat#8.region_cat	-
449	15.ethnic_cat#10.region_cat	-
450	15.ethnic_cat#11.region_cat	-
451	16.ethnic_cat#2.region_cat	-
452	16.ethnic_cat#3.region_cat	-
453	16.ethnic_cat#4.region_cat	-
454	16.ethnic_cat#5.region_cat	-
455	16.ethnic_cat#8.region_cat	-
456	16.ethnic_cat#9.region_cat	-
457	16.ethnic_cat#10.region_cat	-
458	17.ethnic_cat#3.region_cat	-
459	17.ethnic_cat#4.region_cat	-
460	17.ethnic_cat#5.region_cat	-
461	17.ethnic_cat#7.region_cat	-
462	17.ethnic_cat#9.region_cat	-
463	17.ethnic_cat#10.region_cat	-
464	17.ethnic_cat#11.region_cat	-
465	18.ethnic_cat#2.region_cat	-
466	18.ethnic_cat#4.region_cat	-
467	18.ethnic_cat#6.region_cat	-
468	18.ethnic_cat#8.region_cat	-
469	18.ethnic_cat#10.region_cat	-
470	18.ethnic_cat#11.region_cat	-
471	19.ethnic_cat#3.region_cat	-
472	19.ethnic_cat#4.region_cat	-
473	19.ethnic_cat#5.region_cat	-
474	19.ethnic_cat#8.region_cat	-
475	19.ethnic_cat#9.region_cat	-
476	19.ethnic_cat#10.region_cat	-
477	19.ethnic_cat#11.region_cat	-

Table R4. List of fixed effects included in MVMR

No.	Fixed effects	Description
1	2.ethnic_cat	African
2	3.ethnic_cat	Any other Asian background
3	4.ethnic_cat	Any other Black background
4	5.ethnic_cat	Any other mixed background
5	6.ethnic_cat	Any other white background
6	7.ethnic_cat	Bangladeshi
7	8.ethnic_cat	British
8	9.ethnic_cat	Caribbean
9	10.ethnic_cat	Chinese
10	11.ethnic_cat	Indian
11	12.ethnic_cat	Irish
12	13.ethnic_cat	Other ethnic group
13	14.ethnic_cat	Pakistani
14	15.ethnic_cat	Mixed
15	16.ethnic_cat	White
16	17.ethnic_cat	White and Asian
17	18.ethnic_cat	White and Black African
18	19.ethnic_cat	White and Black Caribbean
19	2.region_cat	East of England
20	3.region_cat	London
21	4.region_cat	North East
22	5.region_cat	North West
23	6.region_cat	Scotland
24	7.region_cat	South East
25	8.region_cat	South West
26	9.region_cat	Wales
27	10.region_cat	West Midlands
28	11.region_cat	Yorkshire and The Humber

Response to Reviewer Comments (Round 3)

Before formally addressing your insightful and constructive comments, we would like to thank you for your time and energy devoted to reviewing this manuscript. Your comments and suggestions have undoubtedly helped improve the manuscript, again. Below, we address your comments in a point-by-point manner; if we may, we have numbered each point of your comments, for ease of addressing. For ease of review, we have quoted your comments in *italic* type and outlined our responses in a standard font.

Reviewer #2 (Remarks to the Author):

Thank you for your efforts and background information on HDFE, as well as the analysis script. The new method is more clearly justified and in general, the paper reads better. The methods are more transparent. I have to say that as I have not used STATA nor this MR package, I am not fully able to tell, what exactly is going on. But experts with those packages can now more easily track how you achieved these results!

The MR script makes it very clear, what covariates are used. These variables do not fully line up with the R4 table, there is education, and townsend index, and less ethnic dummies than in table R4. Could you please fix the table R4 to reflect what was done in the script?

Finally, as I said, I found the additional material supplied to me very useful. I think some of your readers may also have similar questions - could you please add the tables R2, R3, and R4 to the supplement and also link to the github repo in the paper? Thanks!

Response: Thank you for your thoughtful and constructive feedback. We have carefully revised the manuscript to ensure consistency between the variables listed in Table R4 (i.e., List of Fixed Effects included in MVMR) and those used in the analysis script. As per your suggestion, we have also included Tables R2, R3, and R4 in the Supplementary Information to provide additional transparency for our readers (i.e., Supplementary Tables 3, 5, and 8 in the new version). Additionally, we have linked to the GitHub repository in the manuscript, enabling readers to access the analysis scripts and further explore our methods. We sincerely appreciate your insightful comments and your effort in reviewing our work, which have greatly improved the clarity and transparency of the manuscript.