**Supplementary Materials**

**Triangulated Evidence Provides No Support for Bidirectional Causal Pathways Between Diet/Physical Activity and Depression/Anxiety**

Kirsten J.M. van Hooijdonk1, Zoe E. Reed2,3, Nina van den Broek1, Madhurbain Singh4,5, Hannah M. Sallis3,6, Nathan A. Gillespie4,7, Marcus R. Munafò2,3, Jacqueline M. Vink1

1 Behavioural Science Institute, Radboud University, Nijmegen, The Netherlands 2 School of Psychological Science, University of Bristol, Bristol, UK
3 MRC Integrative Epidemiology Unit, University of Bristol, Bristol, UK
4 Virginia Institute for Psychiatric and Behavioral Genetics, Department of Psychiatry, Virginia Commonwealth University, Richmond VA, USA
5 Department of Human and Molecular Genetics, Virginia Commonwealth University, Richmond VA, USA
6 Centre for Academic Mental Health, Population Health Sciences, Bristol Medical School, University of Bristol, UK 7 QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia
Correspondence to: kirsten.vanhooijdonk@ru.nl. Postal address: Postbus 9104, 6500 HE, Nijmegen, The Netherlands

|  |  |
| --- | --- |
| **Content** | Page  |
| **A.** Deviations From Preregistration | 2 |
| **B.** Study 1 Procedure and Participants | 3 |
| **C.** Study 1 Details on Measures | 4 |
| **D.** Study 2 Details on Measures | 7 |
| **E.** Study 2 Details on Analyses | 9 |
| **F.** Study 3 Details per Phenotype | 10 |
| **G.** Study 3 Selection of Genetic Instruments | 12 |
| **H.** Study 3 Performed Sensitivity Analyses | 13 |

**Supplementary Material A**. **Deviations From Preregistration.**

***Study 1: Random Intercept Cross-Lagged Panel Models***

* We did add the calculation of interclass corrections (ICCs) for sweet snack intake, savoury snack intake, fruit intake, vegetable intake, physical activity, depressive symptoms, and anxiety symptoms, as this provides information on the proportion of variance explained by between-person variance and within-person variance across the waves.
* The continuous total physical activity scores were scaled for use in the mixed-effects models. As default the continuous physical activity scores were divided by 100. However, some models still did not converge. Therefore in those models, physical activity scores were divided by 1000, which solved the convergence issues.
* In the preregistration, we planned to standardise the food intake variables allowing for comparison of the different food intake categories. However, as food intake variables were not modelled simultaneously, we used the raw scores instead of the standardised scores.

***Study 2: Co-Twin Control Design Mixed-Effects Models***

* The continuous total physical activity scores were scaled for use in the mixed-effects models by dividing the physical activity scores by 100.

***Study 3: Mendelian Randomization***

* UVMR and MVMR: For the phenotype Major Depressive Disorder (MDD), only the summary statistics of UK Biobank (UKBB) and the Psychiatric Genomics Consortium (PGC) were used, as the summary statistics of 23andMe\_307k were not publicly available and the available summary statistics file from the meta-analyses including all three datasets (with 10,000 variants) didn’t include the SNPs of our interest.
* UVMR and MVMR: MR-GRAPPLE [1] was not included as a sensitivity analysis in both UVMR and MVMR because MR-GRAPPLE requires two nonoverlapping full summary statistics files for the exposure (one used for SNP selection and one used for estimating the marginal effects of the SNPs on this exposure). As this was not available for the traits of our interest, we decided to exclude MR-GRAPPLE from the planned list of sensitivity analyses. For UVMR, we replaced MR-GRAPPLE for MR-RAPS as an alternative to correct for bias due to weak instruments [2]. For MVMR, no suitable alternative was available and therefore MR-GRAPPLE was excluded from the list of MVMR sensitivity analyses.
* MVMR: Eight MVMR models with physical activity as the outcome were added to the planned set of MVMR models mentioned in the preregistration. These models were already planned to be conducted for the other two methods included in the triangulation framework but were by mistake missing in the MR preregistration.

**References**

1. Wang Y, Zhao R, Wang B, Zhao C, Zhu B, Tian X. The Dose-Response Associations of Sugar-Sweetened Beverage Intake with the Risk of Stroke, Depression, Cancer, and Cause-Specific Mortality: A Systematic Review and Meta-Analysis of Prospective Studies. Nutrients. 2022;14(4). <https://doi.org/10.3390/nu14040777>

2. Zhao Q, Wang J, Hemani G, Bowden J, Small DS. Statistical inference in two-sample summary-data Mendelian randomization using robust adjusted profile score. Ann. Statist. 2018. <https://doi.org/10.1214/19-AOS1866>

**Supplementary Material B.** **Study 1 Procedure and Participants.**

***Study 1: Random Intercept Cross-Lagged Panel Model***

The Healthy Student Life project is an ongoing study in which Dutch university students from Radboud University are followed over time. During each wave, student complete a questionnaire assessing their mental health and lifestyle. In this study, data from the first three waves were used. Wave 1 was collected October-November 2021 (N invited = 25,035), Wave 2 was collected May-July 2022 (N invited = 23,994) and Wave 3 was collected May-July 2023 (N invited = 23,425). During each wave, all students enrolled at the university at that moment were invited by mail to complete an online questionnaire. Recruitment activities (e.g., social media and email reminders) were used to increase the number of responses, and incentives were raffled off among all participants. The initial total sample consisted of 9,978 students. However, for use in the current study, 702 students were excluded because only background information from the administrative systems of the university was available and no questionnaire data. This resulted in an analytical sample of n = 9,276 students, with 6,004 students participating in one questionnaire, 2,208 students participating in two questionnaires and 1,064 students participating in all three questionnaires. The study has been independently reviewed by Radboud University's Social Sciences Ethics Committee, and there was no formal objection to this research (ECSW-2021-086).

**Supplementary Material C.** **Study 1 Measures.**

***Study 1: Random Intercept Cross-Lagged Panel Model***

**1. Sweet snack intake.** To assess participants’ sweet snack intake we asked: “The following questions are about your eating and snacking habits. Think about a normal week and indicate how many days per week (0-7) you eat … 1) Pastry, cake and/or large cookies (e.g., a piece of pie, slice of cake, donut, muffin, stuffed biscuit, caramel waffle, pink cake, brownie, tompouce), 2) Candybars (e.g., Nuts, Mars, KitKat, Snickers, B’tween with chocolate, Balisto), 3) Chocolate (e.g., chocolate bar, deluxe chocolates), and 4) Candy (e.g., liquorice, marshmallow, sour candy, yoghurt gums, wine gums, lollipops)” [1]. Answer options included: (0) 0 days, (1) 1 day, (2) 2 days, (3) 3 days, (4) 4 days, (5) 5 days, (6) 6 days or (7) 7 days. Next, the sum of the number of days per week of the four categories was calculated. The sum theoretically ranged between 0 and 28. Higher scores indicate more days per week that sweet snacks were consumed. In case participants were missing data on one or more of the items, the sum was coded as missing.

 **2. Savoury snack intake.** To assess participants’ savoury snack intake we asked: “The following questions are about your eating and snacking habits. Think about a normal week and indicate how many days per week (0-7) you eat … 1) Potato crisps (e.g., Lay’s, Pringles, Doritos, Croky) and 2) Warm and/or fried snacks (e.g., French fries, croquette, frikandel, sausage roll/bread, spring roll, steamed bun, kipcorn, slice of pizza)” [1]. Answer options included: (0) 0 days, (1) 1 day, (2) 2 days, (3) 3 days, (4) 4 days, (5) 5 days, (6) 6 days or (7) 7 days. Next, the sum of the number of days per week of the two categories was calculated. The sum theoretically ranged between 0 and 14. Higher scores indicate more days per week that savoury snacks were consumed. In case participants missed data on one or more of the items, the sum was coded as missing.

**3. Fruit intake.** To assess participants’ fruit intake we asked: “The following questions are about your eating and snacking habits. Think about a normal week and indicate how many days per week (0-7) you eat … 1) Fruit (e.g., apple, banana, melon, grapes)” [1]. Answer options included: (0) 0 days, (1) 1 day, (2) 2 days, (3) 3 days, (4) 4 days, (5) 5 days, (6) 6 days or (7) 7 days. The score theoretically ranged between 0 and 7. Higher scores indicate more days per week fruits that were consumed.

**4. Vegetable intake.** To assess participants’ vegetable intake we asked: “The following questions are about your eating and snacking habits. Think about a normal week and indicate how many days per week (0-7) you eat … 1) Salad and/or raw vegetables (e.g., lettuce, arugula, cherry tomatoes, slice of cucumber, small carrots, bell pepper) and 2) Cooked, baked, steamed or otherwise heated vegetables (e.g., broccoli, peas, cauliflower, endive, green beans, mushrooms, beetroot, zucchini)” [1]. Answer options included: (0) 0 days, (1) 1 day, (2) 2 days, (3) 3 days, (4) 4 days, (5) 5 days, (6) 6 days or (7) 7 days. Next, the sum of the number of days per week of the two categories was calculated. The sum theoretically ranged between 0 and 14. Higher scores indicate more days per week that vegetables were consumed. In case participants missed data on one or more of the items, the sum was coded as missing.

**5. Physical activity.** Total physical activity was assessed using the International Physical Activity Questionnaire – short form (IPAQ-SF) [2]. This questionnaire asks the number of days and time per day an individual engaged in 1) vigorous physical activity, 2) moderate physical activity and 3) walking. The IPAQ-SF scoring guide was used to calculate a continuous total MET/minutes per week physical activity score of vigourous physical activity, moderate physical activity and walking.

Several data cleaning steps were applied: 1) In case the reported time per day on an activity was lower than 10 minutes, the time per day for that activity was recoded to zero. 2) In case the total time of vigorous physical activity, moderate physical activity and walking per day exceeded 960 minutes (16 hours; assuming 8 hours of sleep), no total MET/minutes per week physical activity score of that participant was calculated. 3) In case the total time per activity category was exceeding 180 minutes (3 hours) per day, data for that specific category was truncated to 180 minutes.

For use in the Random Intercept Cross-Lagged Panel Models (RI-CLPMs), the continuous total physical activity scores were scaled by dividing the scores by 100. If this was not sufficient and convergence warnings were encountered, scores were divided by 1000.

**6. Depressive symptoms.** Depressive symptoms were assessed using the 8-item Center for Epidemiological Studies – Depression Scale (CES-D-8) [3]. This questionnaire consists of 8 items and assesses how much of the time during the past week individuals felt e.g., depressed. Answer options included: (0) None or almost none of the time, (1) Some of the time, (2) Most of the time, or (3) All or almost all of the time. A sum score of all items was calculated where higher scores indicate more depressive symptoms. The sum theoretically ranged between 0 and 24. Before calculating the sum score, positively framed items were reverse coded. A sum score was only calculated for individuals with ≥80% of the items completed. When <20% of the items were missing the sum was calculated based on the mean of the completed items. For individuals who had >20% of the items missing, the sum was coded as missing. Cronbach’s alpha was .85 for all waves.

**7. Anxiety symptoms.** Anxiety symptoms were assessed using the 2-item Generalized Anxiety Disorder scale (GAD-2) [4]. This questionnaire consists of 2 items and assesses how often over the last two weeks individuals have been bothered by the following problems: 1) Feeling nervous, anxious or on edge and 2) Not being able to stop or control worrying. Answer options included: (0) Not at all, (1) Several days, (2) More than half the days or (3) Nearly every day. A sum score of these two items was calculated where higher scores indicate more anxiety symptoms. The sum theoretically ranged between 0 and 6. In case participants missed data on one or more of the items, the sum was coded as missing. Cronbach’s alpha was .80 for W1, .82 for W2 and .79 for W3.

**8. Socio-demographics.** Socio-demographics included: a) age, b) gender, c) living situation, d) relationship status, e) educational type of father, f) educational type of mother, g) Body Mass Index, h) perceived physical health and i) perceived mental health.

***Age.*** For descriptive purposes, age (in years) per wave was calculated based on date of birth and date of participation. For the RI-CLPMs, age at W1 was used. However, when age at W1 was not available (for example because participants joined the study at a later wave), their age at W1 was estimated based on the available data on date of birth/date of participation at a later wave.

***Gender.*** Participants were asked, “What is your gender”. Answer options included: (1) Male, (2) Female or (3) Other. In case inconsistencies in reporting gender across the waves were observed, data on gender of this participant across all waves was set to missing (*nset to missing W1 = 43, nset to missing W2 = 39, nset to missing W1 = 34)*. For descriptive purposes, data on all three answer categories and all three waves were used. For the RI-CLPMs, gender at W1 was used. However, we only used categories (1) Male and (2) Female as the number of participants who answered gender (3) Other was small (*n* varied between 42 and 61 across waves). Therefore, in the RI-CLPMs, gender (3) Other data was set to missing. When gender at W1 was not available (for example because participants joined the study at a later wave), gender at a follow-up wave was used instead.

***Living situation.*** Participants were asked, “With whom do you live (multiple answers possible)? Note. Think about your living situation during most of an average work week”. Participants could select one or more of the following categories: “With no one, I live alone”, “Roommate(s) (no family)”, “Parent(s)”, “Partner”, “Child/children”, “Other family member(s)” or “Other, namely…”. Per category, participants were coded as (1) Yes (when they selected that specific category) and (0) No (when that specific category was not selected but at least one of the other living situation categories was selected).

***Relationship status.*** Participants were asked, “Which of the following options best describes your current relationship status?”. Answer options included: (1) Single, (2) Married, (3) In a steady relationship and living together, (4) In a steady relationship and living apart, (5) Dating without a steady relationship, (6) In a non-monogamous relationship, (7) Divorced or (8) Other, namely.

***Educational type father.*** Participants were asked to report the highest completed level of education of their biological father. Answer options included: (1) Less than secondary education, (2) Secondary education, (3) Higher education or (4) Do not know.

***Educational type mother.*** Participants were asked to report the highest completed level of education of their biological mother. Answer options included: (1) Less than secondary education, (2) Secondary education, (3) Higher education or (4) Do not know.

***Body Mass Index.*** Participants were asked, “What is your height in cm?” and “What is your current weight in kg?”. For participants who participated in multiple waves, height was only asked once at the first time of participation. Body Mass Index (BMI) was calculated using the answers on weight and height (weight in kg divided by height in m2). BMI values below 10 and above 60 were set to missing as deemed non-viable.

***Overall perceived physical health.*** Participants were asked, “In general, would you say your physical health is…” [5]. Answer options included: (1) Poor, (2) Fair, (3) Good, (4) Very good or (5) Excellent.

***Overall perceived mental health.*** Participants were asked, “In general, would you say your mental health is…” [5]. Answer options included: (1) Poor, (2) Fair, (3) Good, (4) Very good or (5) Excellent.

**References**

1. van den Broek N, Larsen JK, Verhagen M, Burk WJ, Vink JM. Is Adolescents' Food Intake Associated with Exposure to the Food Intake of Their Mothers and Best Friends? Nutrients. 2020;12(3). <https://doi.org/10.3390/nu12030786>

2. IPAQ Research Committee. Guidelines for data processing and analysis of the International Physical Activity Questionnaire (IPAQ)-short and long forms. 2005. <https://sites.google.com/view/ipaq/score>.

3. Van de Velde S, Levecque K, Bracke P. Measurement equivalence of the CES-D 8 in the general population in Belgium: a gender perspective. Archives of Public Health. 2009;67:1-15. <https://doi.org/10.1186/0778-7367-67-1-15>

4. Donker T, van Straten A, Marks I, Cuijpers P. Quick and easy self-rating of Generalized Anxiety Disorder: validity of the Dutch web-based GAD-7, GAD-2 and GAD-SI. Psychiatry Res. 2011;188(1):58-64. <https://doi.org/10.1016/j.psychres.2011.01.016>

5. DeSalvo KB, Fisher WP, Tran K, Bloser N, Merrill W, Peabody J. Assessing measurement properties of two single-item general health measures. Qual Life Res. 2006;15(2):191-201. <https://doi.org/10.1007/s11136-005-0887-2>

**Supplementary Material D. Study 2 Measures.**

***Study 2: Co-Twin Control Design Mixed-Effects Models***

**1. Twin information.** For all twins, zygosity status was available: 1) Monozygotic (MZ), 2) Dizygotic (DZ) same-sex or 3) DZ opposite-sex.

**2. Sweet/Savoury snack intake.** Sweet/savoury snack intake was assessed using the question: “Over the past week, how often have you consumed: Energy-dense snacks (e.g. confectionary, cakes, sweet biscuits, potato crisps)?” [1]. Answer options included: (1) Once a week, (2) A few times a week, (3) Most days, (4) Once a day, (5) Several times a day, (6) Not at all or (7) Don’t know.

For the main analyses, a dichotomous variable was constructed and individuals were classified as (0) “unexposed” (original categories 1 and 6) and (1) “exposed” (original categories 2-5) to sweet/savoury snacks.

For the sensitivity analyses, a dichotomous variable was constructed and individuals were classified as (0) “unexposed” (original categories 1, 2 and 6) and (1) “exposed” (original categories 3-5) to sweet/savoury snacks.

**3. Fruit/Vegetable intake.** Fruit/vegetable intake was assessed using the question: “Over the past week, how often have you consumed: Fruit and vegetables (either canned, fresh, frozen or dried)?” [1]. Answer options included: (1) Once a week, (2) A few times a week, (3) Most days, (4) Once a day, (5) Several times a day, (6) Not at all or (7) Don’t know.

For the main analyses, a dichotomous variable was constructed and individuals were classified as (0) “unexposed” (original categories 4 and 5) and (1) “exposed” (original categories 1-3 and 6) to insufficient fruit/vegetable intake.

For the sensitivity analyses, a dichotomous variable was constructed and individuals were classified as (0) “unexposed” (original categories 3-5) and (1) “exposed” (original categories 1, 2 and 6) to insufficient fruit/vegetable intake.

**4. Physical activity.** Total physical activity was assessed using the International Physical Activity Questionnaire – short form (IPAQ-SF) [2]. This questionnaire assessed the number of days and time per day an individual engaged in vigorous physical activity, moderate physical activity and walking. The IPAQ-SF scoring guide was used to calculate a total MET/minutes per week physical activity score of vigour physical activity, moderate physical activity and walking. Several data cleaning steps were applied: 1) In case the reported time per day on an activity was lower than 10 minutes, the time per day for that activity was recoded to zero. 2) In case the total time of vigorous physical activity, moderate physical activity and walking per day exceeded 960 minutes (16 hours; assuming 8 hours of sleep), no total MET/minutes per week physical activity score of that participant was calculated. 3) In case the total time per activity category was exceeding 180 minutes (3 hours) per day, data for that specific category was truncated to 180 minutes. For use in the mixed-effects models, the continuous total physical activity scores were scaled by dividing the scores by 100. This score was used in the mixed-effects models when physical activity as the outcome or second exposure variable.

Next, the continuous total MET/minutes per week physical activity score was used to classify participants' total physical activity level as (1) Low, (2) Moderate, or (3) High. The classification guidelines for the IPAQ guideline were used accordingly [2]. Afterwards, a dichotomous variable was constructed which was used for the main analyses and individuals were classified as (0) “unexposed” (when physical activity was moderate or high) and (1) “exposed” (when physical activity was low) to physical inactivity. This dichotomous variable was used in the mixed-effects models when physical activity was used as the exposure on which dis/concordance was determined.

For the sensitivity analyses, a dichotomous variable was constructed and individuals were classified as (0) “unexposed” (when physical activity was high) and (1) “exposed” (when physical activity was low or moderate) to physical inactivity.

**5. Depressive/anxiety symptoms.** Depressive/anxiety symptoms were assessed using the Kessler 10 (K10) scale [3]. This questionnaire consists of 10 items and assesses how often in the past four weeks individuals felt e.g., tired out for no good reason. Answer options included: (1) None of the time, (2) A little of the time, (3) Some of the time, (4) Most of the time or (5) Almost all of the time. A sum score of all items was created (range 10-50) where higher scores indicate more symptoms of depression/anxiety. This continuous score was used in the mixed-effects models when symptoms of depression/anxiety was the outcome

Next, this continuous score was used to classify participants as (1) Low (sum score 10-15), (2) Moderate (sum score 16-21), (3) High (sum score 22-29) and (4) Very high levels (sum score 30-50) of psychological distress. The guidelines by the Australian Bureau of Statistics (ABS) were used accordingly [4].

For the main analyses, a dichotomous variable was created and individuals were classified as (0) “unexposed” (category low psychological distress) and (1) “exposed” (category moderate, high and very high psychological distress) to depressive/anxiety symptoms. This dichotomous variable was used in the mixed-effects models when symptoms of depression/anxiety were used as the exposure on which dis/concordance was determined.

For the sensitivity analyses, a dichotomous variable was created and individuals were classified as (0) “unexposed” (category low and moderate psychological distress) and (1) “exposed” (category high and very high psychological distress) to depressive/anxiety symptoms.

**6. Socio-demographics.** Socio-demographics included: a) age, b) biological sex, c) living situation, d) study/work status, e) highest level of completed education, f) current relationship status, g) Body Mass Index, h) overall perceived physical health and i) overall perceived mental health.

***Age.*** Participants were asked, “How old are you?”.

***Biological sex.*** Participants were asked, “What is your biological sex?”. Answer options included: (1) Female or (2) Male.

***Living situation.*** Participants were asked, “Who do you live with?”. Answer options included: (1) Alone, (2) Alone with children, (3) Partner and no children, (4) Partner and children, (5) Parents, (6) Other relatives, (7) Friends, (8) Shared accommodation or (9) Other.

***Study/work status.*** Participants were asked, “Which of these best describes your main activity currently?”. Answer options included: (1) Student attending school, (2) Student attending university, Technical and Further Education (TAFE) or other education, (3) Full-time work greater than or equal to 30 hours paid employment per week, (4) Part-time work less than 30 hours paid employment per week, (5) Apprenticeship / Traineeship, (6) Volunteer work, (7) Unemployed / looking for work, (8) Home duties, (9) Have a job, but not at work due to illness, vacation, etc, (10) Not working and currently receiving sickness allowance or disability support pension or (11) Other.

***Highest level of education.*** Participants were asked, “What is your highest level of education?”. Answer options included: (1) No formal education, (2) Completed or partially completed primary school (up to grade 7), (3) Completed or partially completed junior high school (grades 8-10), (4) Completed or partially completed senior high school (grades 11-12), (5) Completed or partially completed Certificate or Diploma (includes TAFE, trade qualification), (6) Completed or partially completed Degree or (7) Completed or partially completed Post Graduate Diploma, Masters or PhD.

***Current relationship status.*** Participants were asked, “Which one of these best describes your current relationship status?”. Answer options included: (1) Single (and have never been married), (2) Married or living with partner, (3) In a relationship but not living with partner (boyfriend/girlfriend), (4) Separated (but still legally married), (5) Divorced or (6) Widowed.

***Body Mass Index.*** Participants were asked, “ How tall are you?” (in cms) and “How much do you weigh?” (in kgs). Body Mass Index (BMI) was calculated using the answers on weight and height (weight in kg divided by height in m2). BMI values below 10 and above 60 were set to missing as deemed non-viable.

***Overall perceived physical health.*** Participants were asked, “How would you rate your overall physical health?”. Answer options included: (1) Very good, (2) Good, (3) Moderate, (4) Bad or (5) Very bad.

***Overall perceived mental health.*** Participants were asked, “How would you rate your overall mental health?”. Answer options included: (1) Very good, (2) Good, (3) Moderate, (4) Bad or (5) Very bad.

**References**

1. Mitchell BL, Campos AI, Rentería ME, Parker R, Sullivan L, McAloney K, Couvy-Duchesne B, Medland SE, Gillespie NA, Scott J. Twenty-five and up (25Up) study: a new wave of the Brisbane Longitudinal Twin Study. Twin Research and Human Genetics. 2019;22(3):154-63. <https://doi.org/10.1017/thg.2019.27>

2. IPAQ Research Committee. Guidelines for data processing and analysis of the International Physical Activity Questionnaire (IPAQ)-short and long forms. 2005. <https://sites.google.com/view/ipaq/score>.

3. Andrews G, Slade T. Interpreting scores on the Kessler Psychological Distress Scale (K10). Aust N Z J Public Health. 2001;25(6):494-7. <https://doi.org/10.1111/j.1467-842x.2001.tb00310.x>

4. Australian Bureau of Statistics. Information paper: use of the Kessler Psychological Distress scale in ABS health surveys. Australian Bureau of Statistics Canberra; 2001.

**Supplementary Material E. Study 2 Details on Analyses.**

***Study 2: Co-Twin Control Design Mixed-Effects Models***

The following additional steps were performed. 1) Sub-models 1 (population-level) were corrected for age (standardized) and gender. 2) In the linear mixed-effects models, *p*-values were determined by computing Type-2 Wald *F*-tests with Kenward-Roger approximation for degrees of freedom as implemented in the Anova function of the package car [1]. 3) In the logistic mixed-effects models, *p-*values were directly obtained via the glmer function (using Wald *Z*-tests) in the lme4 package [2]. 4) For all models, confidence intervals were estimated using the confint function (95%, method = bootstrapping, number of simulations = 1000) of the stats package [3].

**References**

1. Fox J, Weisberg S. An R companion to applied regression, Third edition: Sage, Thousand Oaks CA; 2019.

2. Bates D, Maechler M, Bolker B, Walker S, Christensen RHB, Singmann H, Dai B, Scheipl F, Grothendieck G. Fitting Linear Mixed-Effects Models Using lme4. Journal of Statistical Software. 2009;67(1):1-48. <https://doi.org/10.18637/jss.v067.i01>

3. R Core Team. R: A language and environment for statistical computing. 2023. <https://www.R-project.org/>.

**Supplementary Material F. Study 3 Details per Phenotype.**

***Study 3: Mendelian Randomization***

**1. Sweet snack intake (N GWAS = 64,949).** For this trait, UK Biobank (UKBB) data analysed as part of the Medical Research Council Integrative Epidemiology Unit (MRC IEU) OpenGWAS project was used [1, 2]. Sweet snack intake was defined as yes (vs. no) and was based on the question “Did you eat any biscuits, chocolate or sweets yesterday? Also includes sweets, cereal bars, chocolate covered raisins, sweet popcorn and other sweet snacks” in a 24-hour dietary recall of the previous day (UKBB field ID: 102250).

**2. Savoury snack intake (N GWAS = 64,949).** For this trait, UKBB data analysed as part of the MRC IEU OpenGWAS project was used [1, 2]. Savoury snack intake was defined as yes (vs. no) and was based on the question “Did you eat any crisps, nuts or savoury snacks yesterday? Also includes seeds and olives.” in a 24-hour dietary recall of the previous day (UKBB field ID: 102400).

**3. Fruit intake (N GWAS = 447,401).** For this trait, the genome-wide association study (GWAS) by Cole et al. [3] was used which used data from the UKBB. Participants were asked while considering their intake over the last year “About how many pieces of FRESH fruit would you eat per DAY? (Count one apple, one banana, 10 grapes etc as one piece; put '0' if you do not eat any)” (UKBB field ID: 1309).

**4. Vegetable intake (N GWAS = 64,949).** For this trait, UKBB data analysed as part of the MRC IEU OpenGWAS project was used [1, 2]. Vegetable intake was defined as yes (vs. no) and was based on the question “Did you eat any beans, lentils, potatoes or vegetables yesterday? Include fresh, tinned, frozen, dried; chips, salad, coleslaw, baked beans, chickpeas, veg in stews, pies etc.” in a 24-hour dietary recall of the previous day (UKBB field ID: 103990).

**5. Physical activity – self-reported Moderate-to-Vigorous Physical Activity
(N GWAS = 377,234).** For this trait, the GWAS by Klimentidis et al. [4] was used which used data from the UKBB on self-reported levels of physical activity. The sum of metabolic equivalent (MET)-minutes/week scores of moderate (MPV) and vigorous physical activity (VPA) has been used to assess self-reported Moderate-to-Vigorous Physical Activity (MVPA). This score was calculated by combining different questions (based on the International Physical Activity Questionnaire – Short Form (IPAQ-SF) by IPAQ Research Committee [5]) on both MPA and VPA. MPA: “In a typical WEEK, how many days did you do 10 min or more of moderate physical activities like carrying light loads, cycling at normal pace? (Do not include walking)” (UKBB field ID: 904). VPA: “In a typical WEEK, how many days did you do 10 min or more of vigorous physical activity? (There are activities that make you sweat or breathe hard such as fast cycling, aerobics, heavy lifting)” (UKBB field ID: 884). For both questions, in case participants reported 1 day or more, participants were asked “How many minutes did you usually spend doing moderate/vigorous activities on a typical DAY” (UKBB field ids: 894 & 914). Next, the following formulas were used by Klimentidis et al. [4] to calculate an MVPA MET-minutes/week score:

The ‘4’ and ‘8’ in the above-mentioned formulas refer to the metabolic equivalents of both MPA and VPA [6].

**6. Depression (N GWAS = 500,199).** For this phenotype, the publicly available summary statistics of the GWAS by Howard et al. [7] were used including three depression-related phenotypes (broad depression, probable major depressive disorder (MDD), and International Classification of Disease (ICD, version 9 or 10)-coded MDD). The GWAS by Howard et al. [7] includes data from the Psychiatric Genomics Consortium (PGC) [8], UKBB [9] and 23andMe. The latter was excluded from the current study as only 10,000 genetic variants were publicly available.

In the PGC cohort, lifetime MDD cases were classified using the international consensus criteria (DSM-IV, ICD-9, or ICD-10) using structured diagnostic instruments from assessments by trained interviewers, clinician-administered checklists, or medical record review [8]. Controls in most samples were screened for the absence of lifetime MDD and selected randomly for the population.

In the UKBB, the case and control status of broad depression was defined by using participants’ self-reported answers to the questions “Have you ever seen a general practitioner for nerves, anxiety, tension or depression?” (UKBB field ID: 2090) or “Have you ever seen a psychiatrist for nerves, anxiety, tension or depression?” (UKBB field ID: 2100). If answering “yes” to one of these questions (at initial assessment visit or at any repeat assessment visit) or if a hospital admission record including a primary or secondary diagnosis of a depressive mood disorder from linked hospital records was available (UKBB field IDs: 41202 and 41204; ICD codes: F32—Single Episode Depression, F33—Recurrent Depression, F34—Persistent mood disorders, F38—Other mood disorders and F39—Unspecified mood disorders), respondents were classified as a case. If answered “no” to one of these questions (in all participated assessments), respondents were classified as controls. Participants with known diagnoses of bipolar disorder, schizophrenia, personality disorder or prescribed antipsychotic medications were excluded.

**7. Anxiety Disorder (AD) (N GWAS = 17,310).** Nine samples representing seven independent cohorts participating in the Anxiety NeuroGenetics Study (ANGST) were included in the meta-analyses by Otowa et al. [10]. These cohorts include (1) Molecular Genetics of Schizophrenia (MGS) controls, (2) PsyCoLaus, (3) Rotterdam Study (RS), (4) Study of Health In Pomerania (SHIP), (5) Queensland Institute of Medical Research (QIMR), (6) TRacking Adolescent’s Individual Lives Survey (TRAILS) and, (7) The NEtherlands STudy of Depression and Anxiety/Netherlands Twin Registry (NESDA/NTR). For the current study, summary statistic data on the case-control approach were used. Respondents were assigned to the Anxiety Disorder (AD) case group in case they met any criteria for lifetime AD (i.e., Generalized AD, panic disorder, social phobia, agoraphobia, and specific phobias) and respondents were assigned to the “super-normal” control group when subjects had few or no clinical anxiety symptoms. Respondents with subsyndromal AD were excluded.

**References**

1. Elsworth B, Lyon M, Alexander T, Liu Y, Matthews P, Hallett J, Bates P, Palmer T, Haberland V, Smith GD, Zheng J, Haycock P, Gaunt TR, Hemani G. The MRC IEU OpenGWAS data infrastructure. bioRxiv. 2020:2020.08.10.244293. <https://doi.org/10.1101/2020.08.10.244293>

2. Elsworth B, Mitchell R, Raistrick C, Paternoster L, Hemani G, Gaunt T. MRC IEU UK Biobank GWAS pipeline version 1. 2017. [https://doi.org/https://doi.org/10.5523/bris.2fahpksont1zi26xosyamqo8rr](https://doi.org/https%3A//doi.org/10.5523/bris.2fahpksont1zi26xosyamqo8rr)

3. Cole JB, Florez JC, Hirschhorn JN. Comprehensive genomic analysis of dietary habits in UK Biobank identifies hundreds of genetic associations. Nat Commun. 2020;11(1):1467. <https://doi.org/10.1038/s41467-020-15193-0>

4. Klimentidis YC, Raichlen DA, Bea J, Garcia DO, Wineinger NE, Mandarino LJ, Alexander GE, Chen Z, Going SB. Genome-wide association study of habitual physical activity in over 377,000 UK Biobank participants identifies multiple variants including CADM2 and APOE. Int J Obes (Lond). 2018;42(6):1161-76. <https://doi.org/10.1038/s41366-018-0120-3>

5. IPAQ Research Committee. Guidelines for data processing and analysis of the International Physical Activity Questionnaire (IPAQ)-short and long forms. 2005. <https://sites.google.com/view/ipaq/score>.

6. Craig CL, Marshall AL, Sjöström M, Bauman AE, Booth ML, Ainsworth BE, Pratt M, Ekelund U, Yngve A, Sallis JF, Oja P. International physical activity questionnaire: 12-country reliability and validity. Med Sci Sports Exerc. 2003;35(8):1381-95. <https://doi.org/10.1249/01.Mss.0000078924.61453.Fb>

7. Howard DM, Adams MJ, Clarke TK, Hafferty JD, Gibson J, Shirali M, et al. Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. Nat Neurosci. 2019;22(3):343-52. <https://doi.org/10.1038/s41593-018-0326-7>

8. Wray NR, Ripke S, Mattheisen M, Trzaskowski M, Byrne EM, Abdellaoui A, et al. Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. Nat Genet. 2018;50(5):668-81. <https://doi.org/10.1038/s41588-018-0090-3>

9. Howard DM, Adams MJ, Shirali M, Clarke TK, Marioni RE, Davies G, Coleman JRI, Alloza C, Shen X, Barbu MC, Wigmore EM, Gibson J, Hagenaars SP, Lewis CM, Ward J, Smith DJ, Sullivan PF, Haley CS, Breen G, Deary IJ, McIntosh AM. Genome-wide association study of depression phenotypes in UK Biobank identifies variants in excitatory synaptic pathways. Nat Commun. 2018;9(1):1470. <https://doi.org/10.1038/s41467-018-03819-3>

10. Otowa T, Hek K, Lee M, Byrne EM, Mirza SS, Nivard MG, et al. Meta-analysis of genome-wide association studies of anxiety disorders. Mol Psychiatry. 2016;21(10):1391-9. <https://doi.org/10.1038/mp.2015.197>

**Supplementary Material G. Study 3 Selection of genetic instruments.**

***Study 3: Mendelian Randomization***

**Univariable Mendelian Randomization.** In all analyses, the SNPs associated with the exposure of interest were identified in a GWAS of that specific exposure. For this selection, the genome-wide significance level (*p* < 5 x 10-8) was used. In the analyses with sweet snack intake, savoury snack intake, vegetable intake or anxiety as exposure, a less stringent threshold was used (*p* < 1 x 10-5) as no or only one exposure SNP(s) were/was available. Afterwards, SNPs were clumped to ensure independence using the“clump\_data” function from the TwoSampleMR package (r2 < .001 and linkage disequilibrium window of 10000 kb) [1] and the exposure SNPs were extracted from a second GWAS of the outcome of interest. In case a SNP was not identified in the outcome GWAS, a proxy SNP was used when available. Proxies were searched for either via the “extract\_outcome\_data” (proxies=TRUE)function in the TwoSampleMR package (for traits where summary statistics were accessed via the IEU Open GWAS project [1]) or via the LDproxy tool from LDlink where the first proxy SNP present in both the exposure and outcome data was selected (for other traits [2]). In both cases, a cut-off of r2 ≥ .8 was used and data from the 1,000 Genomes project was used as the European reference population [3]. Afterwards, exposure and outcome data were harmonized, palindromic SNPs were identified and only included in analyses where it was possible to infer the forward strand alleles. Supplementary Table S6 provides an overview of the SNPs (including proxy SNPs when relevant) included per UVMR model.

**Multivariable Mendelian Randomization.** A similar procedure as described for UVMR was used to select genetic instruments for MVMR. SNPs for exposure 1 and exposure 2 were selected using the same *p-*value thresholds, after which SNPs associated with exposure 1 and 2 were combined in one set of exposure SNPs and clumping was performed. In case not all SNPs were available in the dataset of exposure 1 and 2, a proxy search was performed. In cases where no proxy SNP could be identified, this SNP was excluded from the set of exposure SNPs. Afterwards, the exposure data was harmonized, the set of exposure SNPs was extracted from the outcome GWAS, proxy SNPs were used for missing SNPs in the outcome data (when available), the exposure and outcome data were harmonized, palindromic SNPs were identified and only included in analyses where it was possible to infer the forward strand alleles. Supplementary Table S8 provides an overview of the SNPs (including proxy SNPs when relevant) included per MVMR model.

**References**

1. Hemani G, Zheng J, Elsworth B, Wade KH, Baird D, Haberland V, Laurin C, Burgess S, Bowden J, Langdon R, Tan V, Yarmolinsky J, Shihab H, Timpson N, Evans D, Relton C, Martin R, Smith D, Gaunt T, Haycock P. The MR-Base platform supports systematic causal inference across the human phenome. elife. 2018;7:e34408. <https://doi.org/10.7554/eLife.34408>

2. Machiela MJ, Chanock SJ. LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. Bioinformatics. 2015;31(21):3555-7. <https://doi.org/10.1093/bioinformatics/btv402>

3. The 1000 Genomes Project Consortium. A global reference for human genetic variation. Nature. 2015;526(7571):68-74. <https://doi.org/10.1038/nature15393>

**Supplementary Material H. Study 3 Performed Sensitivity Analyses**

***Study 3: Mendelian Randomization***

**Univariable Mendelian Randomization.** The following sensitivity analyses were performed:

* MR-Egger to correct for directional pleiotropy [1].
* Weighted median which allows relaxation of the instrumental variable assumptions [2].
* Simple mode which provides valid estimates, regardless of the type of horizontal pleiotropy, if the most common horizontal pleiotropy value is zero [3].
* Weighted mode which provides valid estimates as long as the most frequent SNP-effects are contributed by valid genetic instruments [3].
* MR Pleiotropy RESidual Sum and Outlier (MR-PRESSO) to identify outliers and correct for horizontal pleiotropy, using the MR-PRESSO package [4].
* MR using the robust adjusted profile score (MR-RAPS) to account for weak instrument bias, using the mr.raps package [5].
* MRlap to correct for sample overlap, using the MRlap package [6].

**Multivariable Mendelian Randomization.** The following sensitivity analyses were performed:

* MVMR-Egger to correct for unmeasured and measured pleiotropy [7].
* MVMR-PRESSO to identify outliers and correct for horizontal pleiotropy, using the MR-PRESSO package [4].

**References**

1. Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. Int J Epidemiol. 2015;44(2):512-25. <https://doi.org/10.1093/ije/dyv080>

2. Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator. Genet Epidemiol. 2016;40(4):304-14. <https://doi.org/10.1002/gepi.21965>

3. Hartwig FP, Davey Smith G, Bowden J. Robust inference in summary data Mendelian randomization via the zero modal pleiotropy assumption. Int J Epidemiol. 2017;46(6):1985-98. <https://doi.org/10.1093/ije/dyx102>

4. Verbanck M, Chen CY, Neale B, Do R. Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. Nat Genet. 2018;50(5):693-8. <https://doi.org/10.1038/s41588-018-0099-7>

5. Zhao Q, Wang J, Hemani G, Bowden J, Small DS. Statistical inference in two-sample summary-data Mendelian randomization using robust adjusted profile score. Ann. Statist. 2018. <https://doi.org/10.1214/19-AOS1866>

6. Mounier N, Kutalik Z. Bias correction for inverse variance weighting Mendelian randomization. Genet Epidemiol. 2023;47(4):314-31. <https://doi.org/10.1002/gepi.22522>

7. Rees JMB, Wood AM, Burgess S. Extending the MR-Egger method for multivariable Mendelian randomization to correct for both measured and unmeasured pleiotropy. Stat Med. 2017;36(29):4705-18. <https://doi.org/10.1002/sim.7492>