**Supplementary Materials Files**

To: **Association of the newly proposed dietary index for gut microbiota and constipation: National Health and Nutrition Examination Survey**

by Zhuhui Zhang, Chunlu Bi, Runsheng Wu, Mouwen Qu

**Supplementary Method**

***Weighted analysis***

Weighted analysis is a statistical method that assigns different weights to observations in a sample to reflect their importance and representativeness within the overall dataset. This technique enhances the accuracy and efficiency of estimates and allows for a deeper understanding of population characteristics and relationships between variables. By adjusting sample weights, weighted analysis corrects estimation biases, leading to more accurate results. Additionally, it optimizes estimation efficiency by assigning different weights to observations based on their relative importance, ensuring that the sample is representative and leading to more accurate population estimates[1, 2]. Detailed information and methodology regarding NHANES weighting can be found in the NHANES guidelines on their website [2]. Following these guidelines, we incorporated the complex sampling design and mobile examination center sample weights into our study, ensuring that our data sample represents a large and diverse U.S. adult population.

***Multiple imputation***

Multiple imputation is a robust and widely used statistical technique for handling missing data across various fields, including epidemiology, social sciences, and clinical research. The method involves generating several plausible datasets through imputation, conducting separate analyses on each dataset, and pooling the results to produce valid estimates and measures of uncertainty [3, 4]. This technique helps mitigate potential biases and ensures the representativeness of the findings.

The multiple imputation by chained equations (MICE) method, employed in this study, is a flexible approach that sequentially imputes missing values for each variable by developing regression models, conditioned on the observed values of other variables [3, 4]. We created five imputed datasets and combined the results according to Rubin’s Rules, ensuring the robustness and reliability of our conclusions.

**References:**

1.NHANES Survey Methods and Analytic Guidelines. https://wwwn.cdc.gov/nchs/nhanes/analyticguidelines.aspx#estimation-and-weighting-procedures. Accessed 12 Aug 2024

2.NHANES Survey Methods and Analytic Guidelines. https://wwwn.cdc.gov/nchs/nhanes/analyticguidelines.aspx#estimation-and-weighting-procedures. Accessed 12 Aug 2024

3.White IR, Royston P, Wood AM (2011) Multiple imputation using chained equations: Issues and guidance for practice. Stat Med 30:377–399. https://doi.org/10.1002/sim.4067

4.Beesley LJ, Bondarenko I, Elliot MR, et al (2021) Multiple imputation with missing data indicators. Stat Methods Med Res 30:2685–2700. https://doi.org/10.1177/09622802211047346

**Supplementary Table:**

**Table S1. Components and scoring criteria of DI-GM in NHANES.**

**Table S2. Association between DI-GM and constipation of the NHANES 2005-2010 participants after** **multiple imputation**

**Table S3 STROBE Statement Checklist**

**Table S1. Components and scoring criteria of DI-GM in NHANES.**

|  |  |  |
| --- | --- | --- |
| **Components of DI-GM** | **Food items included in NHANES** | **Scoring criteria** |
| **Beneficial to gut microbiota** | Avocados | Score 1 - Consumption≥sex-specific median  Score 0 - Otherwise |
| Broccoli |
| Chickpeas |
| Coffee |
| Cranberries |
| Fermented dairy (including yogurt, cheese, kefir, sour cream, buttermilk) |
| Fiber |
| Soybean (including Soy milk, Tofu) |
| Whole grains |
| **Unfavorable to gut microbiota** | Refined grains | Score 0 - Consumption≥sex-specific median  Score 1 - Otherwise |
| Processed meat |
| Red meat |
| High-fat diet (% energy) | Score 0 - Consumption≥40%  Score 1 - Otherwise |

Abbreviations: DI-GM, dietary index for gut microbiota; NHANES, National Health and Nutrition Examination Survey.

**Table S2. Association between DI-GM and constipation of the NHANES 2005-2010 participants after** **multiple imputation**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Variable** | **Model 1a** | | **Model 2b** | | **Model 3c** | | **Model 4d** | |
| **OR(95% CI)** | **P-value** | **OR(95% CI)** | **P-value** | **OR(95% CI)** | **P-value** | **OR(95% CI)** | **P-value** |
| DI-GM | 0.81 (0.76, 0.86) | <0.001 | 0.84 (0.78, 0.90) | <0.001 | 0.83 (0.77, 0.89) | <0.001 | 0.83 (0.77, 0.89) | <0.001 |
| DI-GM groupe |  |  |  |  |  |  |  |  |
| 0-3 | 1 (Reference) |  | 1 (Reference) |  | 1 (Reference) |  | 1 (Reference) |  |
| 4 | 0.76 ( 0.59, 0.99) | 0.041 | 0.83 (0.63, 1.08) | 0.164 | 0.80 (0.61, 1.06) | 0.115 | 0.80 (0.61, 1.06) | 0.116 |
| 5 | 0.58 ( 0.44, 0.76) | <0.001 | 0.65 (0.49, 0.87) | 0.005 | 0.62 (0.46, 0.84) | 0.003 | 0.62 (0.46, 0.84) | 0.003 |
| ≥6 | 0.44(0.33,0.58) | <0.001 | 0.52 (0.38, 0.71) | <0.001 | 0.50 (0.36, 0.68) | <0.001 | 0.50 (0.37, 0.69) | <0.001 |
| Tend test |  | <0.001 |  | <0.001 |  | <0.001 |  | <0.001 |
| Beneficial to gut microbiota | 0.70 (0.65, 0.76) | <0.001 | 0.76 (0.69, 0.83) | <0.001 | 0.76 (0.69, 0.85) | <0.001 | 0.77 (0.69, 0.86) | <0.001 |
| Unfavorable to gut microbiota | 1.01 (0.92, 1.12) | 0.768 | 0.99 (0.90, 1.09) | 0.804 | 0.92 (0.83, 1.01) | 0.083 | 0.91 (0.82, 1.00) | 0.057 |

Abbreviations: DI-GM, dietary index for gut microbiota; OR, odd ratio; CI, confidence interval;

a Model 1: unadjusted for any covariates.

b Model 2: adjusted for age+sex+race+marital status+PIR+education level.

c Model 3: Model2+smoking status+drinking status+physical activity time+BMI+total calories intake.

d Model 4: Model3+CVD history+Hypertension+DM+depression.

e..The DI-GM ranges from 0-11 (including beneficial to gut microbiota [ranges from 0-7] and unfavorable to gut microbiota [ranges from 0-4]) and grouped according to 0-3, 4, 5, and ≥6

**Table S3 STROBE Statement Checklist**

**Item**

**No Recommendation**

**Page No**

|  |  |
| --- | --- |
| **Title and abstract** 1 (*a*) Indicate the study’s design with a commonly used term in the title or the  abstract  (*b*) Provide in the abstract an informative and balanced summary of what was done and what was found | 1  1 |

**Introduction**

|  |  |  |  |
| --- | --- | --- | --- |
| Background/rationale | 2 | Explain the scientific background and rationale for the investigation being reported | 2 |
| Objectives | 3 | State specific objectives, including any prespecified hypotheses | 2 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Methods**  Study design | 4 Present key elements of study design early in the paper | | 2-3 | |
| Setting | 5 | Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection | 3 | |
| Participants | 6 | (*a*) Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up  (*b*) For matched studies, give matching criteria and number of exposed and unexposed | 3 | |
| Variables | 7 | Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable | 3-4 | |
| Data sources/ measurement | 8\* | For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group | 3-4 | |
| Bias | 9 | Describe any efforts to address potential sources of bias | NA | |
| Study size | 10 | Explain how the study size was arrived at | NA | |
| Quantitative variables | 11 | Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why | 5 | |
| Statistical methods | 12 | (*a*) Describe all statistical methods, including those used to control for confounding  (*b*) Describe any methods used to examine subgroups and interactions  (*c*) Explain how missing data were addressed  (*d*) If applicable, explain how loss to follow-up was addressed (*e*) Describe any sensitivity analyses | 5-6  5-6  6 | |
| **Results** |  |  |  | |
| Participants | 13\* | (a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study,  completing follow-up, and analysed  (b) Give reasons for non-participation at each stage  (c) Consider use of a flow diagram | 3  3  3 | |
| Descriptive data | 14\* | (a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders  (b) Indicate number of participants with missing data for each variable of interest | 6  3 | |
| (c) Summarise follow-up time (eg, average and total amount) | | | | NA | |
| Outcome data 15\* Report numbers of outcome events or summary measures overtime | | | | 6 | |

|  |  |  |  |
| --- | --- | --- | --- |
| Main results | 16 | (*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included  (*b*) Report category boundaries when continuous variables were categorized  (*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period | 6-7  4 |
| Other analyses | 17 | Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses | 7 |
| **Discussion** | | | |
| Key results | 18 | Summarise key results with reference to study objectives | 7 |
| Limitations | 19 | Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias | 8 |
| Interpretation | 20 | Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence | 7-8 |
| Generalisability | 21 | Discuss the generalisability (external validity) of the study results | 8 |
| **Other information** | | | |
| Funding | 22 | Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based | No  funding |

\*Give information separately for exposed and unexposed groups.

**Note:** An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting.