Additional file 1 Supplementary methods

Comparing performance of three propensity score weighting methods for continuous exposures in nutritional epidemiology

Yuriko Muramatsu, Tosiya Sato, Hirohito Sone, Shiro Tanaka

This supplementary material includes the following contents.

S1 Dataset

We used an anonymized dataset derived from the Japan Diabetes Complications Study (JDCS). The JDCS was originally designed as an open-label, randomized controlled trial and was later continued as a cohort study and followed up participants to evaluate the association between lifestyle factors and complications of type 2 diabetes [1]. The median follow-up period was 7.8 years [2]. Between January 1995 and March 1996, 2,205 participants were enrolled at outpatient clinics at 59 universities or general hospitals specializing in diabetes care in Japan, and 2,033 participants were eligible for the previous study. The JDCS dataset includes demographic data at baseline, nutrient intake measured in 1996 and 2000, annual clinical test results measured during the follow-up period, and incidence of complications.

S2 Energy adjustment

Various discussions have been made on the role of total energy intake in estimating the association between nutrients and health outcomes. It has been reported as a confounding factor, as total energy intake is frequently associated with nutrients and serves as a risk factor for diseases [3]. Another study considered total energy intake as a collider, given that it is produced from exposure and other nutrients [4]. By adjusting for energy, the exposure effect represents the effect of substituting exposure intake with other nutrients' intake while keeping the total energy intake at the same level [4]. This reflects the research question in nutritional epidemiology studies on how altering the composition of the diets without changing their absolute amount affects health-related outcomes [5, 6]. In this study, total energy intake was treated as the intermediate variable since total dietary fiber produces a little energy. Therefore, the effect of the adjusted total dietary fiber intake follows the effect of isocaloric substitution, as described above. We suggest that the role of total energy intake be defined according to the hypotheses of each study.

This study adjusted total dietary fiber intake using the residual method [7]. Firstly, total dietary fiber intake was square-root transformed to approximate the normal distribution. Secondly, the square-root-transformed values were regressed on the total energy intake using a linear regression model, and the residuals between the intake estimates based on the dietary assessment and those from the regression model were calculated. Finally, energy-adjusted total dietary fiber intake was obtained by adding the expected total dietary fiber intake to the residuals.

S3 Imputation of missing covariates

Missing values in covariates were imputed once using fully conditional specification, with linear regression for continuous variables and logistic regression for binary variables. The regression models included 15 baseline covariates in the generalized propensity score model, the incidence of stroke and CHD, and other baseline covariates measured in the JDCS, including diastolic blood pressure, treatment with lipid-lowering agents, treatment with antihypertensive agents, estimated glomerular filtration rate, waist circumference, waist-to-hip ratio, albumin-to-creatinine ratio, sedentary job, and fasting plasma glucose.

S4 Assumptions

In this study, we assumed consistency, positivity, conditional exchangeability, and no-interference.

Consistency assumes that the outcome of each participant we observe is equal to the potential outcome they would have had if they had received the exposure which was actually received [8].

The positivity assumption requires that the probability of receiving each level of exposure is not zero for any combination of covariate and exposure levels [9]. Specifically, we assumed that there was no situation in which a subgroup stratified by covariate and exposure levels included no participants who could receive exposure due to contraindications [9]. In contrast, we assumed that the practical violation of positivity might occur when some subgroups did not include exposed participants because of the limited sample size [9].

Conditional exchangeability assumes that exposure and potential outcomes are independent of each other, conditional on confounding factors [8].

We assumed no interference in this study. Interference refers to a situation in which the potential outcome of a participant is influenced by the exposure received by another participant [8].

S5 Quantile binning approach and generalized propensity score-based weights

In quantile binning approach, to rank nutrient intakes into bins, we utilized the SAS command "proc rank" and assigned the largest of the corresponding ranks to the tied data values.

The notations and mathematical formulas used to create the weights are described below. The covariate vector of participant i is defined as X_i $(i = 1, \dots, n)$. The probability that the nutrient intake of participant i belongs to bin j is defined as $a_j(X_i)$ $(j = 1, \dots, J)$ and is calculated using the multinomial logistic regression models shown below. Here β and γ stand for an intercept and a vector of regression coefficients of covariates, respectively.

$$a_1(X_i) = \frac{1}{1+\sum_{k=2}^{J}\exp(\beta_k + X_i^t\gamma_k)} \quad (j = 1),\ a_j(X_i) = \frac{\exp(\beta_j + X_i^t\gamma_j)}{1+\sum_{k=2}^{J}\exp(\beta_k + X_i^t\gamma_k)} \quad (j = 2, \dots, J) \qquad \text{(equation 1)}$$

Using the estimated generalized propensity score, each weight was constructed using equation 2 to 4.

For IPTW weights [10], $\frac{1}{J \times a_j(X_i)}$ (equation 2)

For GOW weights [11], $\frac{1}{a_j(X_i) \times \sum_{k=1}^{J}\frac{1}{a_k(X_i)}}$ (equation 3)

For GMW weights [12], $\frac{\min(a_1(X_i),\dots,a_J(X_i))}{a_j(X_i)}$ (equation 4)

S6 Marginal standardization and percentile bootstrap methods

As the reference for the estimates in the entire study population, which is the target population of IPTW, marginal standardization using logistic regression models was used [13, 14]. In detail, the expected incidence risks of stroke or CHD were estimated as if each participant had the average nutrient intake, or as if each participant had a 1-unit increase in the intake from the average intake. The expected incidence risks were then averaged across all participants. The risk ratio was calculated as the ratio of the mean expected risk of stroke or CHD in the case of 1-unit increase in intake from the average to the mean expected risk in the case of the average intake.

To evaluate the precision of the estimated risk ratios, 95% confidence intervals were calculated using the percentile bootstrap method. The generalized propensity score was estimated for each of the 1,000 bootstrap samples and then was used to create weights. Using the weighted log binomial regression models, risk ratios were calculated for each bootstrap sample, and the 2.5th and 97.5th percentiles of the risk ratios were selected as the lower and upper limits of the confidence intervals, respectively. When separation occurred in estimating the generalized propensity score using multinomial logistic regression models, or when the calculation of risk ratios did not converge, the corresponding bootstrap samples were removed from contributing to the construct of the percentile bootstrap confidence intervals.

S7 Effective sample size

The effective sample size was used to assess weight variability. We used the following notation and mathematical formulas.

D_{ij} is the indicator variable and is assigned a value of 1 when participant i belongs to bin j, and 0 otherwise. $w_j(X_i)$ stands for the weight of participant i in bin j.

$$\left(\sum_{i=1}^{n}\sum_{j=1}^{J} D_{ij} w_j(X_i)\right)^2 / \sum_{i=1}^{n}\sum_{j=1}^{J} D_{ij} w_j^2(X_i) \qquad \text{(equation 5)}$$

S8 Weighted correlation coefficient

It is suggested that 0.1 can be used as the threshold for the absolute weighted correlation coefficient between exposure and each covariate to interpret whether confounding effect is small [15, 16]. In this study, we judged that the covariate balance improved if the absolute weighted correlation coefficient decreased compared to the unweighted value. We suggest that each study should choose how to evaluate the weighted correlation coefficient. When thresholds are used, they can be decided depending on the covariate variability in the analysis population.

S9 Statistical software

We performed all analyses using SAS 9.4 (SAS Institute Inc., Cary, NC, USA). R version 4.4.1 was used to create figures.

References cited in the supplementary methods

1. Tanaka S, Tanaka S, Iimuro S, et al. Cohort profile: The Japan Diabetes Complications Study: a long-term follow-up of a randomised lifestyle intervention study of type 2 diabetes. Int J Epidemiol. 2014;43(4):1054-62.

2. Sone H, Horikawa C, Tanaka-Mizuno S, et al. Japan Diabetes Complications Study: Revisiting one of the first large-scale clinical studies in East Asians with diabetes. J Diabetes Investig. 2025;16(3):360-9.

3. Willett WC, Howe GR, Kushi LH. Adjustment for total energy intake in epidemiologic studies. Am J Clin Nutr. 1997;65(4 Suppl):1220S-8S; discussion 1229S-31S.

4. Tomova GD, Arnold KF, Gilthorpe MS, et al. Adjustment for energy intake in nutritional research: a causal inference perspective. Am J Clin Nutr. 2022;115(1):189-98.

5. Willett WC, Stampfer M, Tobias DK. Re: Adjustment for energy intake in nutritional research: a causal inference perspective. (editorial). Am J Clin Nutr. 2022;116(2):608-9.
6. McCullough LE, Byrd DA. Total Energy Intake: Implications for Epidemiologic Analyses. Am J Epidemiol. 2023;192(11):1801-5.
7. Willett W, Stampfer MJ. Total energy intake: implications for epidemiologic analyses. Am J Epidemiol. 1986;124(1):17-27.
8. Hernán M, Robins J. Causal Inference: What If: Boca Raton: Chapman & Hall/CRC; 2020. https://content.sph.harvard.edu/wwwhsph/sites/1268/2024/01/hernanrobins_WhatIf_2jan24.pdf. Accessed 14th September 2025.
9. Cole SR, Hernán MA. Constructing inverse probability weights for marginal structural models. Am J Epidemiol. 2008;168(6):656-64.
10. Naimi AI, Moodie EE, Auger N, et al. Constructing inverse probability weights for continuous exposures: a comparison of methods. Epidemiology. 2014;25(2):292-9.
11. Li F, Li F. Propensity score weighting for causal inference with multiple treatments. Ann Appl Stat. 2019;13(4):2389-415.
12. Yoshida K, Hernández-Díaz S, Solomon DH, et al. Matching weights to simultaneously compare three treatment groups: comparison to three-way matching. Epidemiology. 2017;28(3):387-95.
13. Sato T. Confounding and effect modification in epidemiologic studies. Proceedings of the Institute of Statistical Mathematics. 1994;42(1):83-101 (in Japanese).
14. Localio AR, Margolis DJ, Berlin JA. Relative risks and confidence intervals were easily computed indirectly from multivariable logistic regression. J Clin Epidemiol. 2007;60(9):874-82.
15. Zhu Y, Coffman DL, Ghosh D. A boosting algorithm for estimating generalized propensity scores with continuous treatments. J Causal Inference. 2015;3(1):25-40.
16. Austin PC. Assessing covariate balance when using the generalized propensity score with quantitative or continuous exposures. Stat Methods Med Res. 2019;28(5):1365-77.