Statistical Association Between Dietary Quality and Risk of Chronic Diseases

Zhiyu Liu *

College of Arts and Sciences, Georgia State University, Atlanta, 30303, USA

* Corresponding Author Email: zliu38@student.gsu.edu

Abstract. Diet quality is believed to be associated with risks of multiple chronic diseases, yet evidence based on different dietary scoring systems needs to be compared and integrated within a unified framework. Using nationally representative NHANES 2017–March 2020 data, we constructed a composite dietary Z-score to assess its associations with cardiovascular disease, type 2 diabetes and obesity. Individual-level HEI-2015, Mediterranean Diet Score (MDS), and DASH scores were standardized using survey-weighted means and standard deviations under the complex sampling design, then summed to form the composite diet Z-score (Zsum). We fitted design-weighted multivariable logistic regression models in three stages: M1 unadjusted; M2 adjusted: M1 + age, sex, race/ethnicity; M3 extended the M2 specification by adjusting for education, poverty-income ratio, smoking status, alcohol consumption, and physical activity level. The primary exposure was Zsum per +1 SD, supplemented by tertile analyses (Low/Mid/High; Low as reference) and trend tests. Results remained robust after accounting for timing and season. Future longitudinal studies with causal inference are warranted to verify directionality and quantify preventable risk.

Keywords: Diet quality; HEI-2015, Mediterranean Diet Score, DASH, NHANES, logistic regression, CVD, T2D, Obesity.

1. Introduction

Chronic non-communicable diseases, including CVD, type 2 diabetes, cancer, and chronic respiratory disease account for over 50% of global deaths [1]. And it has become a major burden on global public health. These conditions not only contribute to premature death but also lead to disability and magnificant healthcare costs. Diet quality, as a modifiable lifestyle factor, bears a close relation to chronic disease development. The evidence increasingly supports that higher diet quality reduces chronic disease risk and entails significant socioeconomic impacts. For instance, one study has shown, that high-quality dietary patterns may have a protective association with cancer in men [2].In recent years, Numerous epidemiological studies and public databases (NHANES, UK Biobank, GBD) have provided data support for exploring the statistical evidence linking diet quality and the risk of chronic diseases .The concept of dietary quality is no longer limited to the intake of individual nutrients or food types, it has expanded to encompass an overall evaluation of the dietary pattern, including indices like the HEI, the Mediterranean diet, and the DASH pattern, which are widely used for quantifying healthy diet [3-5]. These indices, through a structured scoring system, can comprehensively reflect the consistency between an individual's diet quality and health conditions.The existing research supports a significant inverse correlation between a high-quality dietary pattern (adequately consume vegetables, fruits, lean protein and healthy fats) and risk of chronic diseases. Diet-linked chronic conditions—obesity, coronary heart disease, T2D, various inflammatory disorders, and some cancers—are frequently termed lifestyle diseases and are associated with dietary modification alongside physical inactivity (World Health Organization) [6]. However, different dietary scoring systems may have differences in their predictive efficacy for disease risks, because most of them were developed to reflect compliance with local national dietary guidelines, and were formulated around specific dietary patterns, which calls for more unified statistical comparisons and analyses. However, cross-country comparability remains limited, as most indices were designed based on local dietary guidelines. Therefore, this study aims to utilize public database resources to comprehensively apply multiple dietary quality evaluation tools, and through advanced statistical methods, systematically explore the connections between dietary quality and the risk of chronic diseases such as CVD, T2D and obesity, in order to reveal the public health value of diet intervention and provide a basis for precision nutrition and policy formulation, and to provide theoretical basis for subsequent intervention studies and policy formulation.

2. Materials and methods

2.1. Materials

We used the publicly available National Health and Nutrition Examination Survey (NHANES) database from 2017-March2020 to quantify the association between diet quality and chronic disease risk. The most recent consecutive period (2017-March2020) of NHANES data was selected, and adult respondents aged 20 years and above were chosen as samples. Exclusions encompassed (1) current pregnancy or lactation; (2) samples with missing data on key variables including diet, BMI, and disease history. The database provides detailed Dietary Interview - Total Nutrient Intakes data, Examination Data, Laboratory Data (such as Insulin, Cholesterol), and Questionnaire Data (Diabetes).

2.2. Dietary Quality Assessment

This study will employ three internationally recognized dietary quality indices to comprehensively evaluate the dietary patterns of the participants.

(1) Healthy Eating Index-2015 (HEI-2015): This gauge's consistency with the U.S. Dietary Guidelines on a 0–100 scale built from 13 components: nine adequacy (fruits, vegetables, whole grains, etc.) and four moderations (refined grains, added sugars, etc.). Higher scores correspond to higher diet quality [7].

(2) Mediterranean Diet Score (MDS): MDS assesses the extent to which intake aligns with a traditional Mediterranean-style pattern. This indice typically comprises 9–11 components (e.g., vegetables, fruits, nuts, fish, olive oil, red meat), each scored relative to the sample median (0 or 1 point). Higher totals reflect better diet quality [8].

(3) Dietary Approaches to Stop Hypertension Score (DASH Score): This measure appraises consistency with a DASH-style dietary pattern. Quintile ranking is applied to each food group's intake, and points are awarded accordingly, with the summed total ranging from 0 to 40. Higher totals indicate stronger adherence and better overall diet quality [9].

After deriving individual HEI-2015, MDS, and DASH scores, we removed scale differences by standardizing each score within the analytic sample using the NHANES complex survey design (i.e., design-weighted mean μ and design-weighted standard deviation σ), according to (1):

$$Z = \frac{(\text{raw acore} - \mu)}{\sigma} \tag{1}$$

We then summed the three Z-scores to obtain a composite dietary Z-score (Zsum), with higher scores denoting better overall dietary quality. In the main analyses, Zsum was entered as a continuous predictor in survey-weighted logistic regression models. To describe and probe potential nonlinearity, Zsum was additionally categorized into design-weighted tertiles (Low, Mid, High; Low as reference), and a P for trend was obtained (using tertile medians). Zsum was computed and analyzed only for participants with complete values for all three component scores and complete survey design information.

2.3. Main Chronic Diseases of Focus and Their Definitions

Cardiovascular disease (CVD) is defined as a situation where an individual affirms that they have been diagnosed by a physician with angina, myocardial infarction which is also known as heart attack, or stroke. Type 2 diabetes (T2D) is defined in case any of these conditions are satisfied: the individual self-reports having received a diagnosis from a physician (with gestational diabetes being excluded from this); the level of HbA1c is equal to or greater than 6.5%; the fasting plasma glucose (FPG) is equal to or greater than 7.0 mmol/L; or the person is undergoing treatment with glucose-lowering agents. Obesity is defined when BMI measured during the examination is equal to or greater than 30 kg/m².

2.4. Statistical Analysis

Analyses were carried out using R, drawing chiefly on the survey package. We strictly adhered to the NHANES complex sampling design for weighting and variance estimation: SDMVSTRA as the stratification variable, SDMVPSU as the primary sampling unit, and survey weights prioritized WTDRD1PP (with WTDR2DPP used when WTDRD1PP was unavailable) [10, 11]. Standard errors were computed via Taylor linearization.

(1) Descriptive analyses. Participants were grouped by tertiles of the composite dietary Z-score (T1 = lowest, T2 = middle, T3 = highest). Continuous data are reported as weighted means (SE) and contrasted using linear regression accounting for the survey design. A linear trend was tested by treating the median of each tertile as a continuous predictor. Categorical data are expressed as weighted proportions (SE) and evaluated via the Rao–Scott χ^2 test. (2) Main models (logistic regression, M1–M3). We employed complex-survey–weighted multivariable logistic regression (svyglm, quasibinomial) to investigate the relationship of the composite dietary Z-score (continuous) with three outcomes: CVD, T2D and obesity, which were summarized as odds ratios and 95% confidence intervals. Three hierarchically adjusted specifications were considered: M1: Unadjusted. M2: M1 + age, sex, race/ethnicity. Model 3: M2 + education (<HS; HS; >HS), PIR, and smoking history (never; former; current). alcohol use (never / former / current), and physical activity (MET-minutes/week). Trend analysis: Tertiles of the composite Z-score were assigned their within-tertile medians and entered as a continuous term to obtain P for trend. (3) Incremental time model (M3S). To account for temporal factors, we extended M3 by additionally adjusting for season and survey cycle: Season (RIDEXMON): 1 = Nov–Apr, 2 = May–Oct (modeled as a factor). Survey cycle (SDDSRVYR): modeled as a categorical variable per NHANES coding. All other covariates matched M3; results are reported as ORs (95% CIs).(4) Missing data. Categorical covariates included an explicit “Missing” level to retain participants. For PIR and physical activity (PA_MET), we used single imputation plus missing-indicator: PIR was imputed to the sample median and PA_MET to 0, with corresponding binary indicators denoting missingness. If a categorical covariate collapsed to a single level within the complete-case subset for a given model, it was omitted to avoid non-identifiability.

3. Results

In the weighted analytic sample, the analytic sample included 12,076 participants for the CVD and T2D models and 12,389 participants for the obesity model. We first examined diet quality as a continuous predictor. Using the comprehensive dietary Z-score (Zsum) as a continuous exposure (per +1 SD) in survey-weighted multivariable logistic regression (M3), we found that (Fig. 1):

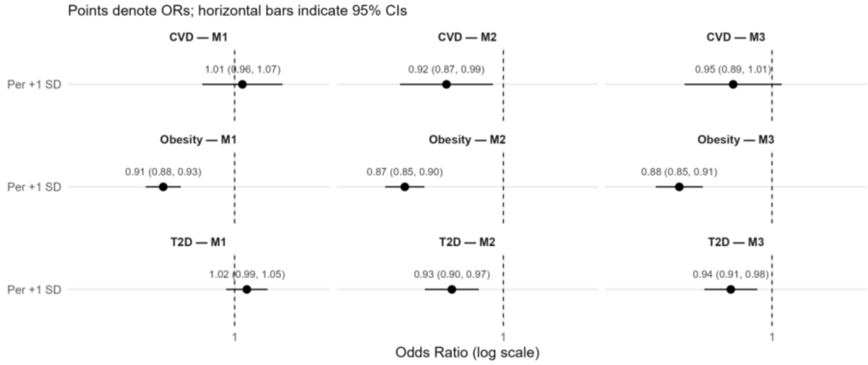

Fig 1. Diet Z-score (per +1 SD) and chronic disease risk: survey-weighted logistic regression (Picture credit: Original)

Obesity risk was notably lower among participants with higher Zsum values (OR = 0.88, 95% CI 0.85–0.91), indicating a robust inverse association between the composite score and adiposity-related outcomes. Similarly, the risk of type 2 diabetes was significantly reduced (OR = 0.94, 95% CI 0.91–0.98), suggesting that metabolic regulation may improve with increasing Zsum levels. In contrast, the risk of cardiovascular disease showed a downward trend (OR = 0.95, 95% CI 0.89–1.01), although the association did not reach statistical significance, implying that the protective effect may be weaker or confounded by other factors (Fig. 2). When Zsum was categorized into tertiles, using the lowest tertile as the reference, a clear and consistent dose–response pattern was observed across the three disease outcomes, further supporting the graded association between higher Zsum and lower cardiometabolic risk.

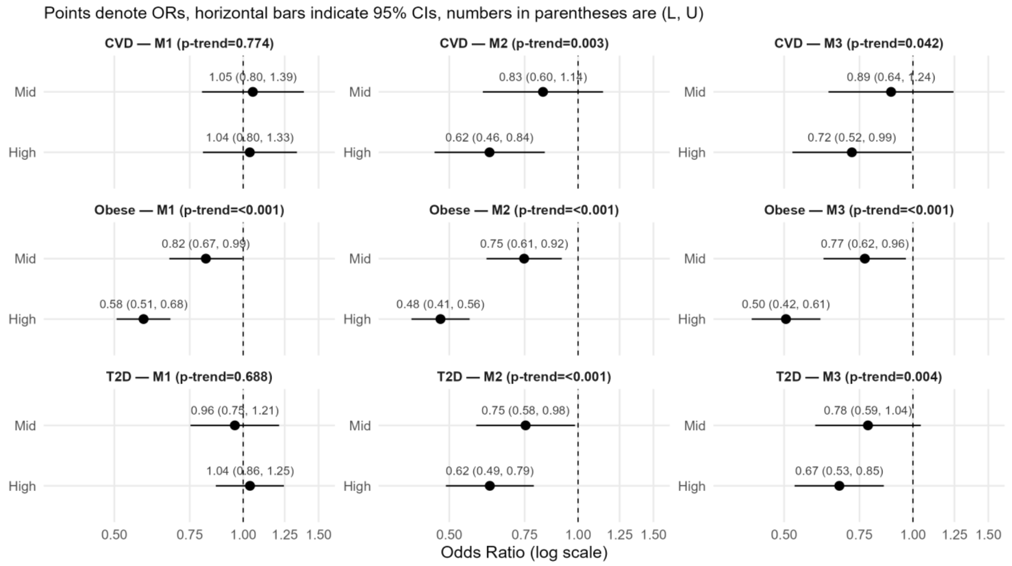

Fig 2. Diet Z-score tertiles (reference = Low) (Picture credit: Original)

In the fully adjusted obesity model (M3), the middle category showed an OR of 0.77 (95% CI: 0.62–0.96) and the highest category an OR of 0.50 (95% CI: 0.42–0.61), with a significant linear trend ($P < 0.001$). In the T2D model (M3), the ORs were 0.78 (95% CI: 0.59–1.04) for the middle category and 0.67 (95% CI: 0.53–0.85) for the highest, with P-trend = 0.004. For CVD (M3), the corresponding ORs were 0.89 (95% CI: 0.64–1.24) and 0.72 (95% CI: 0.52–0.99), with P-trend = 0.042 (Fig. 3).

After additionally adjusting for season (RIDEXMON) and survey cycle (SDDSRVYR) in a sensitivity model (M3S), the conclusions were essentially unchanged: the inverse associations with obesity and T2D remained robust and similar in magnitude to M3, whereas the CVD association was directionally consistent but still not statistically significant. Overall, a higher comprehensive dietary Z-score was consistently linked to lower risks of obesity and T2D, while the protective association with CVD was mainly evident in tertile comparisons and trend analyses.

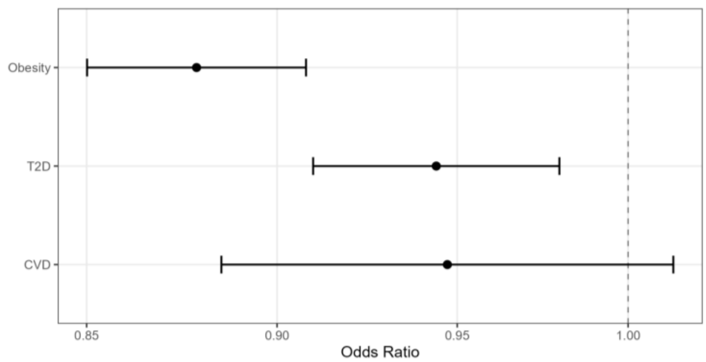

Fig 3. Association between comprehensive Diet Z-score and CVD/T2D/Obesity: M3S (Picture credit: Original)

4. Summary

This study offers evidence at the national level concerning the connection between the quality of one's overall diet and the risk of developing chronic diseases. Using NHANES sample and multivariable logistic regression with complex survey weighting (M3), each 1-SD increment in the composite dietary Z-score was inversely related to obesity (OR 0.88; 95% CI 0.85–0.91), reflecting a 12% reduction in odds as well as having type 2 diabetes (T2D) (the odds were decreased by 6%, with an OR of 0.94 and a 95% CI from 0.91 to 0.98). And there was suggestive, non-significant evidence of protection against CVD (OR 0.95; 95% CI 0.89–1.01). When the Z-score was categorized into tertiles (Low/Mid/High), we observed a consistent dose–response pattern: risks of obesity and T2D were significantly lower in the High group, the CVD analyses (tertile contrasts and trend tests) also supported an inverse association. After additionally adjusting for season and survey cycle (M3S), the conclusions were essentially unchanged, indicating robustness to temporal/seasonal factors.Overall, higher overall diet quality was stably linked to lower risks of obesity and T2D, and may confer some protection against cardiovascular disease. Strengths include constructing a standardized index that integrates HEI-2015, MDS, and DASH and applying rigorous survey weighting; limitations include the cross-sectional design and dietary measurement error. Future studies should employ longitudinal cohorts with repeated dietary assessments, augmented by causal inference frameworks such as marginal structural models, to clarify temporal directionality and quantify the proportion of chronic disease burden potentially preventable through diet improvement.

References

[1] Yach, D., Leeder, S. R., Bell, J., & Kistnasamy, B. (2005). Global chronic diseases. Science, 307(5708), 317. https://doi.org/10.1126/science.1108656

[2] Solbak, N. M., Xu, J. Y., Vena, J. E., Csizmadi, I., Whelan, H. K., & Robson, P. J. (2017). Diet quality is associated with reduced incidence of cancer and self-reported chronic disease: Observations from Alberta's Tomorrow Project. Preventive Medicine, 101, 178–187. https://doi.org/10.1016/j.ypmed.2017.06.009

[3] Guasch-Ferré, M., & Willett, W. C. (2021). The Mediterranean diet and health: A comprehensive overview. Journal of Internal Medicine, 290(3), 549–566. https://doi.org/10.1111/joim.13333

[4] Sacks, F. M., Svetkey, L. P., Vollmer, W. M., Appel, L. J., Bray, G. A., Harsha, D., Obarzanek, E., Conlin, P. R., Miller, E. R., Simons-Morton, D. G., & Karanja, N. (2001). Effects on blood pressure of reduced dietary sodium and the Dietary Approaches to Stop Hypertension (DASH) diet. New England Journal of Medicine, 344(1), 3–10. https://doi.org/10.1056/NEJM200101043440101

[5] Krebs-Smith, S. M., Pannucci, T. E., Subar, A. F., Kirkpatrick, S. I., Lerman, J. L., Tooze, J. A., Wilson, M. M., & Reedy, J. (2018). Update of the Healthy Eating Index: HEI-2015. Journal of the Academy of Nutrition and Dietetics, 118(9), 1591–1602. https://doi.org/10.1016/j.jand.2018.05.021

[6] Medina-Remón, A., Tresserra-Rimbau, A., Pons, A., & Castell, M. (2017). Dietary patterns and the risk of obesity, type 2 diabetes mellitus, cardiovascular diseases, asthma, and neurodegenerative diseases. Critical Reviews in Food Science and Nutrition, 58(2), 262–296. https://doi.org/10.1080/10408398.2016.1158690

[7] Trichopoulou, A., Costacou, T., Bamia, C., & Trichopoulos, D. (2003). Adherence to a Mediterranean diet and survival in a Greek population. New England Journal of Medicine, 348(26), 2599–2608. https://doi.org/10.1056/NEJMoa025039

[8] Fung, T. T., Chiuve, S. E., McCullough, M. L., Rexrode, K. M., Logroscino, G., & Hu, F. B. (2008). Adherence to a DASH-style diet and risk of coronary heart disease and stroke in women. Archives of Internal Medicine, 168(7), 713–720. https://jamanetwork.com/journals/jamainternalmedicine/fullarticle/414155

[9] Zipf, G., Chiappa, M., Porter, K. S., Ostchega, Y., Lewis, B. G., & Dostal, J. (2013). Health and nutrition examination survey: Plan and operations, 1999–2010. National Center for Health Statistics, Vital and Health Statistics, 1(56).

[10] Johnson, C. L., Paulose-Ram, R., Ogden, C. L., Carroll, M. D., Kruszon-Moran, D., Dohrmann, S. M., & Curtin, L. R. (2013). National Health and Nutrition Examination Survey: Analytic guidelines, 1999–2010. Vital and Health Statistics, Series 2, (161).

[11] Rao, J. N. K., & Scott, A. J. (1987). On simple adjustments to chi-square tests with sample survey data. The Annals of Statistics, 15(1), 385–397. https://www.jstor.org/stable/2241089