**Working title:** Dietary Patterns Associated with Food Insecurity Predict a Worse Prognosis for U.S. Cancer Survivors: NHANES 1999-2018

**Introduction**

A cancer diagnosis can upend several facets of life and wellbeing. In addition to the psychological distress that may ensue from the diagnosis itself, both financial toxicity and its accompanying distress can emerge for many cancer survivors as a result of exorbitant treatment and prescription costs and indirect costs [1]. Lost income due to cancer-associated job loss or disability are some examples of the latter [2]. These phenomena we describe are often magnified for low-income cancer survivors, who may lack financial reserves and workplace accommodations while navigating the early treatment phases of their cancer [2,3]. Thus, critical questions arise as to how these experiences can impact cancer-related outcomes in cancer survivors.

Food insecurity, or the lack of continuous access to healthy and nutritious foods to lead a healthy life, can be a consequence faced by cancer survivors with high financial toxicity burdens [4–6]. A framework of competing demands is often used to conceptualize the manifestation of food insecurity amongst cancer survivors [4]. However, of critical public health concern is that food insecurity is associated with negative health outcomes and lower diet quality in the general public [7–9]. Amongst cancer survivors, food insecurity may predict a worse prognosis though more research is needed to substantiate this conjecture.

In a previous analysis, we characterized major dietary patterns describing dietary consumption amongst food insecure cancer survivors using nationally representative data from the National Health and Nutrition Examination Survey (NHANES) [10]. Using penalized logistic regression as a novel supervised learning method for extracting dietary patterns from observed 24-hour recall data, we extracted two major dietary patterns that were strongly and positively associated with food insecurity status amongst cancer survivors. These patterns emphasized consumption of added sugars, processed foods, and deemphasized fruit and vegetable consumption. Concurrently, we found that a “prudent” dietary pattern that emerged and that emphasized healthful dietary components such as fruits, vegetables, and whole grains, was inversely associated with food insecurity status in the same population. The question of whether these dietary patterns impact clinically-meaningful outcomes for cancer survivors was left open-ended.

Understanding how food insecurity impacts different aspects of life, including dietary intake, is a means of delineating at least one of driving factors behind the health disparities that may arise for cancer survivors experiencing food insecurity. Therefore, the goal of this analysis was to examine associations between prevailing dietary patterns in the food insecure cancer survivor population and the risk of mortality among both the broader cancer survivor population and the food-insecure cancer survivor population using nationally representative data. The hypothesis was that the dietary patterns describing consumption patterns in the food insecure cancer survivor population would be positively associated with mortality in the cancer survivor and food insecure cancer survivor populations.

**Methods**

*Study Setting and Population*

We employed data from ten consecutive cycles (1999-2018) from the NHANES, a biennial cross-sectional study implemented by the Centers for Disease Control and Prevention (CDC) and the National Center for Health Statistics that samples civilian and non-institutionalized community dwellers in the United States. The study implements a complex multi-stage sampling design that generates a nationally representative sample. The purpose of the study is to characterize relationships between lifestyle, medical, physiological, and other factors and health outcomes. The study uses surveys that span numerous facets of health and lifestyle and includes a medical examination for a subset of participants. All subjects provided informed and written consent and all study protocols were approved by the NCHS Ethics Review Board [11]. Because this analysis involved deidentified secondary data, it was exempt from Institutional Review Board approval at the University of Illinois Urbana-Champaign.

In Figure 1 we detail the sample flow that arrives at the final analytical sample size of cancer survivors (*n* = 2493), which can be divided into food secure subjects (*n* = 2176) and food insecure subjects (*n* = 317). Food insecurity status was measured using the United States Department of Agriculture’s U.S. Food Security Survey Module (U.S. FSSM) consisting of 18 items designed to evaluate the degree of food insecurity experienced by a subject’s household over the preceding year [12,13]. The survey consists of a series of “yes/no” questions and responses in the affirmative are used to categorize a household as food insecure (responding in the affirmative to ≥ 3 items) or food secure (responding in the affirmative to ≤ 2 items). Cancer status was ascertained via self-reported cancer history on the Medical Conditions Questionnaire (MCQ). We note that individuals with only a diagnosis of non-melanoma skin cancer and no other cancer were coded as having no history of cancer given that the prognosis and benign course of this class of malignancies may otherwise bias the sample [14].

Diagram

Description automatically generated

**Figure 1**. Sample flow diagram detailing inclusion and exclusion criteria for arriving at the final sample.

*Explanatory Variables: Diet Quality Measures*

A thorough description of the computation of the diet quality indices used in the analysis is described elsewhere [10]. We will describe these briefly. Dietary intake data were amassed by NHANES study staff through two 24-hour recalls using the USDA Automated Multiple-Pass Method (for cycles between 1999-2002, only a single 24-hour recall was performed) [15,16]. Nutrient intake data were estimated by referencing the Food and Nutrient Database for Dietary Studies [17]. Dietary intake and nutrient data were averaged across both 24-hour recalls as previously described [10,18,19]. We used the USDA Food Patterns Equivalents Database (FPED) and MyPyramid Equivalents Database (MPED) to obtain intake equivalents of 37 USDA food patterns components and collapsed these further into 26 groups, as previously described [10]. Empirical diet quality measures were extracted from the observed dietary data with penalized logistic regression (penalized logit) and principal components analysis (PCA). The 26 food groups discussed were used as the explanatory variables in these models. In the case of the penalized logit models, four binary outcomes were regressed on the centered and scaled transformations of the explanatory variables and included: food insecurity status (food insecure vs. food secure), age ≥ 60 years, household receipt of SNAP benefits in the last 12 months, and household size ≥ 5, which are all direct measures, surrogate measures, or risk factors of food insecurity [20,21]. See Maino Vieytes et al. for an expanded narrative on these procedures and a discussion about the component retention process for the PCA [10].

*Response Variables: All-Cause and Cause-Specific Mortalities*

Mortality and time-to-event data were acquired from the NHANES Public-Use Linked Mortality File, which is generated from deterministic and probabilistic linkages of the NHANES survey data (through the 2017-2018 cycle) with the National Death Index, described elsewhere [22,23]. We computed time-since-diagnosis and used it as the time scale in our models to minimize the potential for bias by accounting for left-truncation due to delayed enrollment in the study following diagnosis [24–26]. Data were right-censored to either the last known date alive or an administrative censoring date of December 31, 2019. Causes of death were classified by the International Classification of Disease, Tenth Revision (ICD-10) codes. The survival analyses examined all-cause mortality and cause-specific mortality—deaths due to neoplastic malignancy (ICD-10 codes C00-C97), and cardiovascular disease (ICD-10 codes I00-I09, I11, I13, I20-I51, and I60-I69) in our analyses.

*Covariates*

Self-reported demographic and socioeconomic data were obtained from the home interview. Characteristics from the demographic survey (DEMO) included age, sex (*male*/*female*), race and ethnicity (*non-Hispanic White* and *other*), the family income-to-poverty ratio (*< 1.3* or *≥ 1.3*), and household size. We also obtained health insurance status (*covered by health insurance*/*not covered by health insurance*) from the health insurance questionnaire (HIQ/HID—for 1999-2004). Behavioral characteristics included smoking status (*current smoker*—currently smoking every day or some days—, *former smoker*—not currently smoking but with a lifetime history of ≥100 cigarettes—, or *never smoker*—a lifetime history of smoking <100 cigarettes), drinking status (*heavy drinker*— ≥ 14 grams/day for women and ≥ 28 grams/day for men—, *moderate drinker*—0.10-13.9 grams/day for women and 0.10-27.9 g/day for men—, and *abstainer*— < 0.10 grams/day), and physical activity (*measured as weekly MET minutes*) were obtained from the smoking (SMQ) questionnaire, dietary assessment data, and the physical activity questionnaires (PAQ and PAQIAF), respectively [27–29]. Health-related covariates included a Charlson Comorbidity Index score (adapted for NHANES) and body mass index measured during the physical examination (BMI—kilograms/m2) [10,30]. Physical disability was assessed using the 19-item and validated NHANES Activities of Daily Living (ADL) scale, found in the physical functioning questionnaire (PFQ) and whose computation is described in detail elsewhere [31,32]. Cancer-related covariates were obtained from the MCQ and time since cancer diagnosis which was computed as the difference between current age at the time of the survey and age at the first diagnosis of cancer.

*Statistical Analysis*

We assessed the relationships between the diet quality measures and all-cause and cause-specific mortalities using Cox Proportional Hazards models. We implemented a variety of model specifications for the conditional log hazard function to assess the robustness of our results.

The model in equation 1 specifies the diet quality index, , using = 4 dummy variables, , that indicate the subject’s membership in one of the quintiles of the diet pattern index score. In equation 2 we conduct a trend test by assigning the subject the median of their respective quintile and then modeling as a continuous variable (). In equation 3 we specify the diet index as a continuous variable scaled by the standard deviation of the index and in equation 4 we specify the diet index with a basis expansion of basis functions (not shown here) for a natural cubic spline. Models fit using equation 4 used four interior boundary knots () (not shown here). Given that model 3 is nested in model 4, we used the likelihood ratio test to assess for non-linearity [33,34]. Additionally, all models included terms () for covariates (). We fit Cox proportional hazards models to data for the entire sample of cancer survivors (*n*  = 2493) and separately on food insecure cancer survivors (*n* = 317). Covariates in these models included age, sex, race, BMI, household size, family income-to-poverty ratio, education status, health insurance status, alcohol intake, smoking status, calories, weekly MET minutes, and the Charlson Comorbidity Index score, food insecurity status, and the receipt of SNAP benefits [35,36]. To account for the possibility of downwardly biased survival estimates from the contributions of subjects distantly removed from a cancer diagnosis to the risk set, we conducted a sensitivity analysis including only subjects with a primary cancer diagnosis within the five years preceding the date of their interview (*n* = 894). We also considered the NHANES ADL as a covariate given that food security can be associated with physical disability and functional deficit but we did not include it in our primary models given significant missingness in this variable and conducted a separate analysis where we further adjusted for it. All analyses accounted for the complex and probability-based sampling methods of the NHANES study by following the analytical guidelines provided by the NCHS and weighting the analyses accordingly. We used = 0.05 as our threshold level for statistical significance and performed all analyses in R v4.2.2 (The R Foundation, Vienna, Austria). All code and data to reproduce these analyses are public and accessible at: <https://github.com/cmainov/nhanes-fi-ca-mortality>.

**Results**

The analysis included 603,360 person-months of contributions to the risk set with 981 deaths from all causes, 343 cancer deaths, and 235 cardiovascular disease-related deaths. The characteristics of the study sample of cancer survivors stratified on food security status are presented in Table 1. On average, food insecure cancer survivors in this sample were younger than food secure survivors, were more likely to be female, non-White, had a lower educational status, were more likely to live under the poverty line, and were less likely to be covered under health insurance. Food insecure cancer survivors were also more likely to live in a home with five or more individuals, be physically or functionally impaired, identify as a current smoker, were less likely to be heavy drinkers, and had a greater comorbidity burden compared to their food secure counterparts.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Table 1**. Epidemiologic characteristics of the study sample. | | | | |
| **Characteristic** | **Combined Sample** (*n* = 2493) | **Food Insecure** (*n* = 317) | **Food Secure** (*n* = 2176) | ***p*** |
| Age | 62.03 (14.85) | 50.4 (16.46) | 63.32 (14.09) | < 0.01 |
| Sex |  |  |  | < 0.01 |
| Male | 1139 (40.9) | 99 (25.1) | 1040 (42.6) |  |
| Female | 1354 (59.1) | 218 (74.9) | 1136 (57.4) |  |
| Race/Ethnicity |  |  |  | < 0.01 |
| Mexican-American | 174 (2.3) | 51 (8.0) | 123 (1.7) |  |
| Other Hispanic | 133 (2.5) | 40 (7.5) | 93 (1.9) |  |
| Non-Hispanic White | 1730 (86.5) | 156 (70.6) | 1574 (88.3) |  |
| Non-Hispanic Black | 376 (6.2) | 56 (9.2) | 320 (5.9) |  |
| Other/Multiracial | 80 (2.4) | 14 (4.6) | 66 (2.1) |  |
| Education Attained |  |  |  | < 0.01 |
| ≤ High School | 1197 (36.5) | 205 (55.5) | 992 (34.4) |  |
| ≥ Some College | 1296 (63.5) | 112 (44.5) | 1184 (65.6) |  |
| Family Income to Poverty Ratio |  |  |  | < 0.01 |
| < 1.3 | 637 (17.3) | 221 (63.5) | 416 (12.2) |  |
| Health Insurance Status |  |  |  | < 0.01 |
| Insured | 2329 (94.1) | 265 (83.7) | 2064 (95.3) |  |
| Household Size |  |  |  | < 0.01 |
| < 5 Persons | 2274 (92.5) | 247 (78.6) | 2027 (94.1) |  |
| ≥ 5 Persons | 219 (7.5) | 70 (21.4) | 149 (5.9) |  |
| NHANES ADL Score | 22.26 (4.56) | 26.1 (7.34) | 21.93 (4.07) | < 0.01 |
| BMI (kg/m2) | 28.92 (6.61) | 29.82 (7.43) | 28.82 (6.51) | 0.08 |
| Weekly MET Minutes | 2249.04 (4387.81) | 5195.27 (8691.45) | 1923.51 (3462.9) | < 0.01 |
| Daily Caloric Intake (kcal) | 1900.17 (679.88) | 1751 (791.25) | 1916.65 (664.54) | 0.02 |
| Charlson Comorbidity Index | 2.98 (1.35) | 3.36 (1.71) | 2.94 (1.3) | < 0.01 |
| SNAP Assistance |  |  |  | < 0.01 |
| Receiving | 347 (11.2) | 158 (55.6) | 189 (6.3) |  |
| Years Since Diagnosis |  |  |  | 0.62 |
| ≤ 5 years | 894 (32.0) | 119 (30.2) | 775 (32.2) |  |
| > 5 years | 1599 (68.0) | 198 (69.8) | 1401 (67.8) |  |
| Smoking Status |  |  |  | < 0.01 |
| Current | 393 (16.9) | 107 (39.2) | 286 (14.4) |  |
| Former | 1021 (39.4) | 79 (21.9) | 942 (41.3) |  |
| Never | 1079 (43.7) | 131 (38.9) | 948 (44.2) |  |
| Alcohol Use |  |  |  | 0.05 |
| Heavy | 268 (13.8) | 23 (4.6) | 245 (14.9) |  |
| Moderate | 381 (15.9) | 32 (16.1) | 349 (15.9) |  |
| None-drinking | 1844 (70.3) | 262 (79.3) | 1582 (69.3) |  |
| Cause of Death |  |  |  | 0.42 |
| Cancer | 343 (36.4) | 30 (31.9) | 313 (36.7) |  |
| Cardiovascular Dz | 235 (25.6) | 11 (20.4) | 224 (25.9) |  |
| Other | 403 (38.0) | 41 (47.7) | 362 (37.4) |  |
| Percentages may not add to 100% given rounding; *p-*values are from chi-square tests for categorical variables and *t*-tests for continuous variables. | | | | |

In Table 2 we present weighted Pearson correlation coefficients between the six dietary patterns extracted and the individual food groups that comprise them. Within the sample of all cancer survivors, the Food Insecurity pattern was characterized by negative correlations with fruits, vegetables, nuts, and whole grains, a high correlation with added sugars, and a weak-to-moderate positive correlation with meat consumption. The SNAP pattern was negatively correlated with fruit and vegetable categories and positively correlated with added sugar consumption. It was, in many ways, similar to the FI pattern and also shared a high correlation with that pattern (*r* = 0.81). The pattern of correlation coefficients for the household size pattern was also similar to those from the FI and SNAP patterns and shared a moderate correlation with the Food Insecurity (*r* = 0.63) and SNAP (*r* = 0.60) patterns. Finally, the two patterns extracted with PCA, in general, appeared to reflect “prudent” patterns, that emphasized the vegetable categories while de-emphasizing added sugars and were negatively correlated with the FI, SNAP, and Household size patterns. On average, food insecure subjects had a significantly higher scores on the food insecurity and SNAP patterns, with a smaller effect size noted for the household size pattern, and lower scores on the Age, Prudent #1, and Prudent #2 patterns compared to food secure subjects (Table 3).

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Table 2**. Pearson correlation coefficients showing the contributions of each food group to the extracted dietary patterns. Correlations amongst the dietary patterns themselves are included at the bottom of the table. | | | | | | |
| **Pattern** | **Food Insecurity (FI) †** | **Age †** | **Food Assistance (SNAP) †** | **Household Size†** | **Prudent #1 ‡** | **Prudent #2 ‡** |
| **Food Groups** |
| Processed Meats | 0.010 | -0.010 | 0.080 | 0.050 | 0.080 | -0.25 |
| Meats | 0.22 | -0.040 | 0.10 | 0.020 | 0.040 | -0.22 |
| Poultry | 0.020 | -0.28 | -0.020 | 0.23 | -0.050 | 0.23 |
| Seafood—High n-3 | -0.13 | 0.010 | -0.11 | -0.020 | -0.050 | 0.27 |
| Seafood—Low n-3 | -0.12 | 0.050 | -0.030 | -0.11 | -0.050 | 0.030 |
| Eggs | 0.10 | 0.080 | 0.080 | 0.20 | 0.16 | -0.020 |
| Solid Fats | 0.13 | 0.010 | 0.22 | -0.030 | 0.13 | **-0.46** |
| Oils | -0.11 | -0.010 | -0.090 | 0.15 | 0.24 | -0.050 |
| Milk | -0.040 | **0.34** | 0.030 | -0.13 | -0.060 | 0.020 |
| Yogurt | -0.090 | 0.040 | -0.080 | -0.29 | 0.040 | **0.31** |
| Cheese | -0.040 | **-0.36** | 0.080 | -0.15 | **0.31** | **-0.31** |
| Alcohol | -0.17 | -0.24 | -0.30 | -0.080 | **-0.38** | -0.15 |
| Fruit—Other | -0.18 | **0.35** | -0.17 | -0.27 | 0.010 | **0.35** |
| Fruit—Citrus, melons, and berries | -0.21 | 0.18 | -0.20 | **-0.34** | 0.030 | **0.44** |
| Tomatoes | -0.16 | -0.020 | -0.080 | -0.25 | **0.42** | -0.010 |
| Dark-Green Vegetables | -0.22 | -0.18 | **-0.26** | -0.20 | 0.20 | **0.44** |
| Dark-Yellow Vegetables | -0.14 | 0.090 | **-0.31** | -0.030 | 0.10 | **0.39** |
| Other Vegetables | **-0.41** | 0.10 | **-0.52** | **-0.35** | **0.36** | **0.31** |
| Potatoes | **0.38** | 0.21 | 0.11 | 0.090 | 0.13 | -0.13 |
| Other Starchy Vegetables | -0.020 | 0.16 | -0.10 | -0.13 | -0.12 | 0.13 |
| Legumes | 0.030 | -0.26 | 0.23 | 0.25 | 0.030 | -0.13 |
| Soy | -0.090 | -0.11 | -0.22 | 0.21 | 0.070 | 0.20 |
| Refined Grains | -0.060 | -0.11 | 0.21 | 0.10 | 0.10 | **-0.41** |
| Whole Grains | -0.19 | **0.41** | -0.24 | -0.25 | -0.040 | 0.27 |
| Nuts | -0.21 | 0.090 | -0.26 | 0.00 | 0.14 | 0.11 |
| **Added Sugars** | **0.67** | -0.24 | **0.58** | **0.44** | -0.27 | **-0.33** |
| FI | -- |  |  |  |  |  |
| Age | -0.27 | -- |  |  |  |  |
| SNAP | **0.81** | **-0.36** | -- |  |  |  |
| Household Size | **0.63** | **-0.49** | **0.60** | -- |  |  |
| Modified Western | -0.24 | 0.090 | **-0.24** | **-0.29** | -- |  |
| Prudent | **-0.40** | **0.33** | **-0.57** | **-0.39** | 0.16 | -- |
| † Dietary pattern obtained using penalized logistic regression. ‡ Dietary pattern obtained using principal components analysis. Correlation coefficients (*r*) ≥ |0.30| are bolded to ease the identification of notable food groups characterizing the different patterns. This analysis was performed on all cancer survivors (*n* = 2493). | | | | | | |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Table 3**. Means and standard deviations of the extracted dietary patterns across levels of food security status. | | | | |
| **Dietary Pattern** | **Combined Sample**  (*n* = 2493) | **Food Insecure** (*n* = 317) | **Food Secure** (*n* = 2176) | ***p*** |
| Food Security Pattern†  Mean (SD) | -0.03 (0.51) | 0.31 (0.63) | -0.07 (0.48) | < 0.01 |
| Age Pattern†  Mean (SD) | -0.10 (0.61) | -0.26 (0.72) | -0.08 (0.6) | < 0.01 |
| SNAP Pattern†  Mean (SD) | -0.05 (0.84) | 0.44 (0.85) | -0.11 (0.82) | < 0.01 |
| Household Size Pattern†  Mean (SD) | 0.00 (0.18) | 0.08 (0.19) | -0.01 (0.18) | < 0.01 |
| Prudent Pattern #1‡  Mean (SD) | 0.04 (0.62) | -0.05 (0.62) | 0.05 (0.62) | 0.05 |
| Prudent Pattern #2‡  Mean (SD) | -0.04 (1.41) | -0.35 (1.46) | -0.01 (1.4) | < 0.01 |
| *p-*values are for survey-weighted t-tests comparing food secure and insecure survivors.  † Dietary pattern obtained using penalized logistic regression. ‡ Dietary pattern obtained using principal components analysis. | | | | |

In our main analysis and after multivariable adjustment, we found significant associations between the extracted dietary patterns and mortality (Tables 4 and 5). Amongst the sample of all cancer survivors, the highest quintile of the Food Insecurity pattern had a 1.52-fold greater risk of all-cause mortality compared to the lowest quintile and a standard deviation increase in the index score was associated with a 23% increased risk of all-cause mortality. Similarly, the highest quintile of the SNAP pattern score had a 2.17-fold increased risk of all-cause mortality compared to the lowest quintile. A standard deviation increase in the SNAP score was associated with a 1.20-fold greater risk of all-cause mortality. Survival curves and spline curves for these relationships are presented in Figure 2. Amongst food insecure cancer survivors, the parameter estimates were similar albeit they had higher variance. In contrast, there were inverse associations noted for the two “prudent” patterns extracted via PCA. Amongst all cancer survivors, the highest quintile of Prudent pattern #1 had a 46% decreased risk of all-cause mortality compared to the lowest quintile and a 20% decreased risk associated with a standard deviation increase in the score. Within food insecure cancer survivors, the highest quintile of Prudent pattern #2 had a 82% reduction in the risk of all-cause mortality compared to the first quintile with a significant test for trend. When we examined cancer-specific mortality, the parameter estimates amongst all cancer survivors were similar to those for all-cause mortality, particular for the Food Insecurity pattern. However, no results other than an inverse association involving Prudent pattern #1 were statistically significant at the level of = 0.05. Considering cardiovascular disease mortality, the effect sizes were close to the null value and we observed a significant and inverse association between Prudent pattern #1 and the risk of cardiovascular disease-related mortality. Further adjusting for the NHANES ADL score did not significantly alter the results (Supplementary Table 2) despite the loss of a large number of subjects from the risk set. Finally, in our sensitivity analysis that included only subjects with a primary cancer diagnosis within the five years before their study interview (Supplementary Table 3), we found that the association between the Food Insecurity pattern and all-cause mortality became slightly magnified. The SNAP pattern in this analysis also had similar results as what had been observed in the main analysis including all cancer survivors. Notably, relationships between Prudent patterns #1 and #2 and all-cause mortality attenuated towards the null.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Table 4**. Adjusted hazard ratios and 95% confidence intervals for the risks of all-cause and cause-specific mortalities, in relation to the dietary patterns, within the cancer survivor sample (*n* = 2493) | | | | | | | | | |
| **Dietary Pattern**d | ***n*** | **Q1** | **Q2** | **Q3** | **Q4** | **Q5** | ***p***a**trend** | **HR**b**continuous** | ***p***c**non-linear** |
| *All-Cause Mortality* | | | | | | | | | |
| Food Insecurity**†** | 2493 | 1.00 | 0.91 (0.66-1.24) | 0.90 (0.65-1.25) | 1.05 (0.72-1.53) | 1.52 (1.01-2.29)\* | 0.03\* | 1.23 (1.06-1.42)\*\* | 0.63 |
| Age**†** | 2493 | 1.00 | 0.85 (0.57-1.27) | 1.05 (0.67-1.65) | 0.94 (0.64-1.40) | 1.13 (0.74-1.72) | 0.41 | 1.08 (0.94-1.24) | 0.83 |
| Household Size**†** | 2493 | 1.00 | 0.99 (0.75-1.32) | 1.05 (0.76-1.45) | 1.05 (0.76-1.45) | 1.17 (0.74-1.86) | 0.46 | 1.06 (0.91-1.23) | 0.93 |
| Food Assistance (SNAP) **†** | 2493 | 1.00 | 1.21 (0.90-1.62) | 1.05 (0.76-1.44) | 1.25 (0.90-1.73) | 2.17 (1.45-3.24)\*\* | < 0.01\*\* | 1.20 (1.03-1.40)\* | 0.12 |
| Prudent #1**‡** | 2493 | 1.00 | 0.92 (0.68-1.25) | 0.95 (0.69-1.32) | 0.68 (0.46-1.00) | 0.54 (0.36-0.82)\*\* | < 0.01\*\* | 0.80 (0.70-0.92)\*\* | 0.11 |
| Prudent #2**‡** | 2493 | 1.00 | 1.32 (0.88-1.99) | 0.92 (0.61-1.40) | 0.96 (0.63-1.46) | 0.94 (0.61-1.44) | 0.34 | 0.97 (0.86-1.09) | 0.29 |
| *Cancer-Specific Mortality* | | | | | | | | | |
| Food Insecurity**†** | 2493 | 1.00 | 0.83 (0.50-1.37) | 0.93 (0.58-1.49) | 0.93 (0.55-1.58) | 1.49 (0.87-2.57) | 0.13 | 1.23 (0.99-1.51) | 0.64 |
| Age**†** | 2493 | 1.00 | 0.64 (0.37-1.11) | 0.69 (0.36-1.29) | 0.88 (0.53-1.47) | 0.62 (0.35-1.10) | 0.30 | 0.92 (0.76-1.11) | 0.49 |
| Household Size**†** | 2493 | 1.00 | 0.84 (0.49-1.43) | 0.97 (0.62-1.52) | 0.76 (0.48-1.20) | 1.00 (0.53-1.91) | 0.89 | 1.04 (0.86-1.26) | 0.39 |
| Food Assistance (SNAP) **†** | 2493 | 1.00 | 0.87 (0.57-1.34) | 0.98 (0.63-1.51) | 1.02 (0.60-1.71) | 1.80 (0.92-3.51) | 0.12 | 1.17 (0.92-1.48) | 0.25 |
| Prudent #1**‡** | 2493 | 1.00 | 0.86 (0.52-1.44) | 0.91 (0.56-1.48) | 0.54 (0.29-0.98)\* | 0.71 (0.41-1.25) | 0.12 | 0.83 (0.70-0.98)\* | 0.84 |
| Prudent #2**‡** | 2493 | 1.00 | 1.38 (0.75-2.52) | 1.01 (0.55-1.86) | 0.68 (0.35-1.35) | 1.03 (0.54-1.96) | 0.49 | 0.99 (0.82-1.21) | 0.04\* |
| *Cardiovascular Disease Mortality* | | | | | | | | | |
| Food Insecurity**†** | 2493 | 1.00 | 0.84 (0.41-1.72) | 1.08 (0.53-2.20) | 1.33 (0.65-2.74) | 1.05 (0.47-2.37) | 0.51 | 1.16 (0.89-1.50) | 0.84 |
| Age**†** | 2493 | 1.00 | 1.48 (0.69-3.18) | 2.23 (1.01-4.92)\* | 1.67 (0.88-3.18) | 1.72 (0.93-3.18) | 0.22 | 1.19 (0.98-1.43) | 0.79 |
| Household Size**†** | 2493 | 1.00 | 1.00 (0.56-1.80) | 1.73 (0.99-3.02) | 1.67 (0.86-3.27) | 1.20 (0.60-2.41) | 0.15 | 1.13 (0.94-1.37) | 0.25 |
| Food Assistance (SNAP) **†** | 2493 | 1.00 | 1.34 (0.69-2.59) | 1.36 (0.77-2.40) | 1.12 (0.62-2.03) | 2.08 (1.13-3.82)\* | 0.12 | 1.14 (0.91-1.43) | 0.95 |
| Prudent #1**‡** | 2493 | 1.00 | 1.12 (0.60-2.08) | 0.86 (0.40-1.83) | 0.54 (0.28-1.06) | 0.40 (0.18-0.88)\* | < 0.01\*\* | 0.70 (0.57-0.87)\*\* | 0.36 |
| Prudent #2**‡** | 2493 | 1.00 | 0.80 (0.41-1.57) | 1.01 (0.52-1.96) | 0.92 (0.45-1.85) | 1.21 (0.54-2.71) | 0.50 | 1.09 (0.88-1.33) | 0.90 |
| \*\* *p* < 0.01; \* *p* < 0.05  a Test for trend across the quintiles of the dietary exposure. See Equation 2 in the main text.  b Hazard ratio for a standard deviation increase in the dietary exposure. See Equation 3 in the main text.  c Likelihood ratio test *p*-value for natural cubic spline model compared to specifying the model with the scaled dietary exposure. See Equation 4 in the main text.  **†** Dietary pattern obtained using penalized logistic regression.; **‡** Dietary pattern obtained using principal components analysis (PCA).  d All models adjusted for age, sex, race, BMI, household size, family income-to-poverty ratio, education status, health insurance status, receipt of SNAP benefits, food insecurity status, alcohol intake, smoking status, total caloric intake, weekly MET minutes, and the Charlson Comorbidity Index score. | | | | | | | | | |

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Table 5**. Adjusted hazard ratios and 95% confidence intervals for the risks of all-cause and cause-specific mortalities, in relation to the dietary patterns, within the food insecure cancer survivor sample (*n* = 317) | | | | | | | | | |
| **Dietary Pattern**d | ***n*** | **Q1** | **Q2** | **Q3** | **Q4** | **Q5** | ***p***a**trend** | **HR**b**continuous** | ***p***c**non-linear** |
| *All-Cause Mortality* | | | | | | | | | |
| Food Insecurity**†** | 317 | 1.00 | 3.05 (0.86-10.83) | 1.92 (0.50-7.44) | 1.88 (0.63-5.58) | 1.67 (0.44-6.34) | 0.76 | 1.13 (0.74-1.73) | 0.78 |
| Age**†** | 317 | 1.00 | 1.13 (0.34-3.72) | 2.09 (0.81-5.42) | 0.75 (0.26-2.13) | 1.13 (0.46-2.76) | 0.79 | 0.83 (0.63-1.09) | 0.64 |
| Household Size**†** | 317 | 1.00 | 4.01 (1.32-12.24)\* | 2.64 (0.82-8.50) | 4.16 (1.29-13.39)\* | 2.73 (0.80-9.29) | 0.13 | 1.20 (0.81-1.77) | 0.69 |
| Food Assistance (SNAP) **†** | 317 | 1.00 | 2.60 (0.85-7.91) | 2.92 (1.15-7.43)\* | 3.94 (1.41-10.99)\*\* | 2.21 (0.63-7.71) | 0.18 | 1.33 (0.95-1.88) | 0.38 |
| Prudent #1**‡** | 317 | 1.00 | 0.47 (0.20-1.06) | 1.73 (0.74-4.08) | 1.12 (0.33-3.81) | 1.00 (0.32-3.10) | 0.71 | 1.06 (0.71-1.59) | 0.59 |
| Prudent #2**‡** | 317 | 1.00 | 0.53 (0.21-1.32) | 0.43 (0.15-1.27) | 0.18 (0.05-0.62)\*\* | 0.30 (0.08-1.18) | 0.05\* | 0.52 (0.24-1.14) | 0.55 |
| *Cancer-Specific Mortality* | | | | | | | | | |
| Food Insecurity**†** | 317 | 1.00 | 3.09 (0.48-20.01) | 0.28 (0.04-1.72) | 0.75 (0.20-2.88) | 1.44 (0.41-5.00) | 0.97 | 1.34 (0.68-2.63) | 0.66 |
| Age**†** | 317 | 1.00 | 2.03 (0.33-12.50) | 0.25 (0.03-2.27) | 1.49 (0.22-9.87) | 1.20 (0.19-7.38) | 0.74 | 0.89 (0.50-1.57) | 0.93 |
| Household Size**†** | 317 | 1.00 | 1.52 (0.34-6.74) | 2.75 (0.49-15.30) | 0.90 (0.10-7.87) | 2.59 (0.33-20.43) | 0.60 | 1.24 (0.61-2.55) | 0.98 |
| Food Assistance (SNAP) **†** | 317 | 1.00 | 1.16 (0.28-4.86) | 0.86 (0.08-8.72) | 1.95 (0.44-8.60) | 1.26 (0.17-9.48) | 0.57 | 1.23 (0.68-2.23) | 0.97 |
| Prudent #1**‡** | 317 | 1.00 | 0.39 (0.04-3.50) | 0.67 (0.11-3.95) | 0.33 (0.05-2.35) | 1.23 (0.27-5.64) | 0.55 | 1.11 (0.54-2.25) | 0.90 |
| Prudent #2**‡** | 317 | 1.00 | 0.69 (0.14-3.39) | 0.17 (0.02-1.38) | 0.44 (0.07-2.80) | 0.19 (0.02-1.91) | 0.18 | 0.56 (0.24-1.32) | 0.91 |
| *Cardiovascular Disease Mortality*e | | | | | | | | | |
| \*\* *p* < 0.01; \* *p* < 0.05  a Test for trend across the quintiles of the dietary exposure. See Equation 2 in the main text.  b Hazard ratio for a standard deviation increase in the dietary exposure. See Equation 3 in the main text.  c Likelihood ratio test *p*-value for natural cubic spline model compared to specifying the model with the scaled dietary exposure. See Equation 4 in the main text.  **†** Dietary pattern obtained using penalized logistic regression.; **‡** Dietary pattern obtained using principal components analysis (PCA).  d All models adjusted for age, sex, race, BMI, household size, family income-to-poverty ratio, education status, health insurance status, receipt of SNAP benefits, alcohol intake, smoking status, total caloric intake, weekly MET minutes, and the Charlson Comorbidity Index score.  e Survival estimates for cardiovascular disease mortality were not estimable given issues with convergence of the optimizer as a result of the relatively low number of events in the riusk set. | | | | | | | | | |

Diagram

Description automatically generated

**Figure 2**. Relationships between the Food Insecurity (panels A and B) and SNAP (panels C and D) patterns and all-cause mortality in cancer survivors (n = 2493). Adjusted survival curves were generated from models specified with quintile dummy variables and spline curves from expanding the diet quality index using a basis expansion for a natural cubic spline with four interior knots. These models adjusted for age, sex, race, BMI, household size, family income-to-poverty ratio, education status, health insurance status, alcohol intake, smoking status, calories, weekly MET minutes, the Charlson Comorbidity Index score, receipt of SNAP benefits, and food insecurity status.

aDiet pattern scores were normalized prior to plotting

**Discussion**

Using a nationally-representative sample of U.S. cancer survivors, we found that dietary patterns associated with being a food insecure cancer survivor were positively associated with all-cause and cancer-specific mortality after adjusting for several confounders. In a previous analysis, we validated the utility of penalized logistic regression as a novel *a posteriori* method to extract dietary patterns associated with a particular risk factor or condition [10]. As a follow-up analysis, the results we present here demonstrate the clinical value of these dietary patterns and their relationships to cancer-related outcomes such as survival. Of the six dietary patterns that we extracted from the observed 24-hour recall data (four with penalized logit and two with PCA), two of these patterns—the Food Insecurity and SNAP patterns—were robustly and positively associated with all-cause and cancer specific mortalities amongst cancer survivors and a subset of food insecure cancer survivors. There was also evidence that the prudent-style patterns extracted with PCA, that were inversely correlated with food insecurity status, were also inversely associated with all-cause and cancer-specific mortalities, although the strength of the evidence for these patterns was not as strong as for the others mentioned. Moreover, the results we observed were robust after performing multiple sensitivity analyses.

Our findings contribute to a body of evidence highlighting adverse associations between aspects of the food insecurity experience and health outcomes. However, our work is novel in that we focused on cancer survivors, a population that has, overall, received relatively little scrutiny within the broader context of food insecurity, despite that this population may have an elevated risk of experiencing food insecurity [4,37,38]. Several lines of evidence tie food insecurity to an increased comorbidity burden, including increased risks of hypertension, hyperlipidemia, diabetes, and mental health conditions [7,39–41]. Moreover, food insecurity is associated with poor overall health status and recent analyses using the NHANES data demonstrated significant and positive associations between food insecurity status and the risk of all-cause and cardiovascular disease mortality [20,36,40,42,43]. Our analysis complements this previous body of work by demonstrating that dietary intake may be a critical aspect in the pathway between food insecurity and increased mortality and morbidity in cancer survivors.

Across studies, place, populations, and time, diet quality is associated with physiological outcomes that may explain the differential propensity for survival like those observed in our analysis. In a longitudinal sample of older adults from the Health and Retirement Study, higher diet quality, as measured by the HEI-2015, was associated with better lipid and C-reactive protein (CRP) profiles as well as with a decreased likelihood for depression and functional deficits [44]. In another longitudinal sample from the Health, Eating, Activity, and Lifestyle (HEAL) prospective cohort study, higher postdiagnosis HEI-2015 scores were associated with lower CRP levels in a sample of breast-cancer survivors [45]. A nested cross-sectional study from the Multiethnic Cohort Study examined relationships between four *a priori* diet quality indices (AHEI-2010, HEI-2010, aMED, and DASH) and a number of serum carotenoids and biomarkers (leptin, HOMA-IR, glucose, CRP, insulin and triglycerides) and found that higher diet index scores were positively associated with the carotenoid markers and inversely associated with the other biomarkers [46]. Finally, in a cross-sectional analysis of newly diagnosed head and neck squamous cell carcinoma patients, higher diet quality, measured by a “whole foods” dietary pattern extracted using PCA from FFQ data, was inversely associated with several pro-inflammatory cytokines [47]. Thus, the link between diet quality and downstream inflammation may explain our observed results, particularly in the context of cancer where higher inflammatory biomarkers exacerbate disease progression resulting in poor prognosis [48–53].

Our findings have policy implications. As we alluded to in our previous analysis, screening for food insecurity is not a clinical best practice widely implemented in cancer clinics [4,10]. Alongside calls from those also working in this area, we believe our results are grounds for pushing food insecurity and cancer survivorship to the forefront of discussion within major medical organizations. Identifying food insecure patients early on in the cancer care continuum in the setting of oncology clinics can facilitate prompt referral to additional support resources. These include accessing supports through a case manager or social worker that assists the cancer survivor leverage personal and community-level resources or provide referrals to federal nutrition programs [4,54]. The establishment of hospital-based food pantries is another avenue that has shown promise for cancer survivors to access nutritious foods they may otherwise lack access to [49]. Thus, tailoring community and higher-level initiatives that prioritize food support throughout the treatment phase may be critical to mitigating the negative health consequences food insecure cancer survivors may experience secondary to a lack of a steady stream of nutritious foods [56]. Though community-level strategies may yield benefit, particularly in the short-term, more comprehensive systems-level approaches, such as innovative medical insurance models, are needed for more wide-ranging effects on food insecure survivors [4,57,58].

This analysis has several strengths including the nationally representative sample from the NHANES, the use of a validated food insecurity measurement tool, the quality and quantity of covariate data we used to account for potential and known confounders, and the quality of the linked mortality data through the NCHS Linked Mortality Files. Nevertheless, there are limitations to note. First, we used a static measure of dietary intake though we know that patterns of dietary intake are largely dynamic and circumstantial and our analysis was not able to account for any variation in dietary intake across time despite using time-to-event measures that occurred substantially after the measurement instance. In a similar vein, it’s worthy to consider that food insecurity can be a transient phenomenon that subjects recover out of and that this may have occurred for participants in the intervening window between the study visit and the time of the observed event or censorship. An additional consideration with respect to the measurement of dietary intake using 24-hour recalls is that it may be subject to systematic measurement error that we were not able to quantify with the available data. Second, with any analysis of observational data and outside of a rigid set of assumptions, we must conclude that unmeasured or residual confounding cannot be excluded and that no causal interpretations can be made with these results. Third, though we did not account for stress as a confounding variable, given limitations with measures of psychological stress or allostatic load on the NHANES survey and our sample size, we believe it is appropriate to conclude that measures of food insecurity, such as those captured by the USDA FSSM, are likely to be highly correlated with measures of stress. Indeed, the US Household FSSM includes questions designed to capture concern and stress about food insufficiency such as:*“(I/We) worried whether (my/our) food would run out before (I/we) got money to buy more”* [13]. Finally, an important reflection concerning the use of the U.S. Household FSSM is that a measure of household food insecurity may not capture the burden of food insecurity exacted on any one individual within that household.

In summary, we conclude that major and prevailing dietary patterns amongst U.S. food insecure cancer survivors, derived empirically using a novel supervised learning methodology, may deleteriously impact cancer-related outcomes such as all-cause and cause-specific mortality. These patterns, characterized by consumption of added sugars and processed foods with concomitant low consumption of fruits, vegetables, whole grains, and other healthful diet components, and our findings signal an urgent public health challenge demanding innovative policy and community-level solutions. We identify several avenues of future research in this area. One avenue includes developing and evaluating community and individual-level interventions for bolstering food security amongst food insecure cancer survivors throughout the early treatment and cost-prohibitive phases of the cancer-care continuum. A second avenue should focus on piloting interventions for medical provider training on screening for food insecurity in oncology settings. A final avenue of research should extend our work and continue surveillance of dietary intake patterns amongst U.S. food insecure cancer survivors using nationally representative data. Ultimately, advances in such areas will hopefully abate the disparities in health outcomes observed by food insecure cancer survivors that our work and that of others highlights.