International Journal of *Environmental Research and Public Health*

MDPI

Article

Empirical Dietary Patterns Associated with Food Insecurity in U.S. Cancer Survivors: NHANES 1999-2018

Christian A. Maino Vieytes [1], Ruoqing Zhu [2], Francesca Gany [3], Amirah Burton-Obanla [4] and Anna E. Arthur [5,*]

[1] Division of Nutritional Sciences, The University of Illinois at Urbana-Champaign; cam17@illinois.edu
[2] Department of Statistics, The University of Illinois at Urbana-Champaign; rqzhu@illinois.edu
[3] Memorial Sloan Kettering Cancer Center; ganyf@mskcc.org
[4] Division of Nutritional Sciences, The University of Illinois at Urbana-Champaign; amirah2@illinois.edu
[5] Department of Dietetics and Nutrition, The University of Kansas Medical Center; aarthur4@kumc.edu
* Correspondence: aarthur4@kumc.edu ;

Abstract: (1) Background: Food insecurity (FI) is a public health and sociodemographic phenomenon that besets many cancer survivors in the United States. FI in cancer survivors may arise as a consequence of financial toxicity stemming from treatment costs, physical impairment, labor force egress, or a combination of those factors. To our knowledge, an understanding of the dietary intake practices of this population has not been delineated but is imperative for addressing the needs of this vulnerable population.; (2) Methods: Using data from NHANES, 1999-2018, we characterized major dietary patterns in the food insecure cancer survivor population: i. penalized logistic regression (logit) and ii. principal components analysis (PCA). We validated these patterns by examining the association of those patterns with food insecurity in the cancer population.; (3) Results: Four dietary patterns were extracted with penalized logit and two with PCA. In the pattern validation phase, we found several patterns exhibited strong associations with FI. The FI, SNAP, and Household Size patterns (all extracted with penalized logit) harbored the strongest associations and there was evidence of stronger associations in those moderately removed from a cancer diagnosis (2 ≥ and < 6 years since diagnosis).; (4) Conclusions: FI may play an influential role on the dietary intake patterns of cancer survivors in the U.S. The results highlight the relevance of FI screening and monitoring for cancer survivors.

Keywords: cancer; dietary patterns; nutritional epidemiology; food insecurity; survivorship

Citation: Lastname, F.; Lastname, F.; Lastname, F. Title. *Int. J. Environ. Res. Public Health* **2022**, *19*, x. https://doi.org/10.3390/xxxxx

Academic Editor: Firstname Lastname

Received: date
Accepted: date
Published: date

1. Introduction

Food insecurity (FI) is the inability to procure sufficient quantities of safe and nutritious foods that promote the physical, emotional, and psychosocial domains of health and well-being [1]. It is a leading public health issue that affected approximately 13.8 million (10.5 %) U.S. households in 2020 and disproportionately implicates low-income households, single-parent households, communities of color, and those with a recent diagnosis of cancer [1,2]. For many households, experiencing a sudden cancer diagnosis and its side effects may worsen FI status. Increasing treatment costs and side effects attributable to those treatments may prompt lower quality of life (QOL) and physical disability in cancer survivors, which magnify the risks of unemployment and financial sequelae [3,4]. The culmination of these and other known risk factors of FI, including younger age, being less educated, belonging to a marginalized community, and having lower income, may ultimately lead to cancer survivors experiencing FI [5]. Moreover, estimates from non-nationally representative data suggest that the prevalence of FI in the cancer survivor (defined as any person with a history of cancer, from the time of diagnosis to the end of life) population may be higher than the national average and it is unclear if time elapsed since diagnosis plays a role in this phenomenon [6–8].

National guidelines from the WCRF/AICR Third Expert Report have developed recommendations that cancer survivors may implement following a diagnosis. These recommendations include dietary modifications that emphasize the consumption of whole grains, vegetables, and fruit while curtailing the consumption of sugar-sweetened beverages and processed meats, as higher intakes of these foods may be associated with an increased cancer risk and worsen the prognosis [9]. Though following these evidence-based guidelines may improve QOL and disease outcomes, it is unclear how FI impacts cancer survivors' capacity to adhere to those recommendations [10–13]. Ultimately, the combination of treatment-associated sequelae and FI may aggravate nutritional inadequacy in food insecure cancer survivors.

Ascertaining population-specific dietary patterns may reveal critical needs and play a role in developing clinical best practices or food policy targeted at specific at-risk populations. Consequently, the goal of this study was to delineate major dietary intake patterns among food insecure cancer survivors by implementing dietary pattern extraction procedures on nationally representative data from the National Health and Nutrition Examination Survey (NHANES**)**. We implement penalized logistic regression, a novel methodology for dietary patterns analysis embraced by colleagues, and principal components analysis (PCA) to empirically characterize the dietary patterns of our target population [14,15]. We subsequently validate those patterns by examining their relationship to the risk of being food insecure**.** To our knowledge, there are no studies evaluating the dietary patterns of food insecure cancer survivors using nationally representative data and this is the first study to employ NHANES data to analyze empirical dietary intake patterns in cancer survivors with self-reported FI.

2. Materials and Methods

Data from ten consecutive cross sections of the NHANES study, between 1999-2018, were employed for the analysis. The analytical outline and strategy are displayed in Figure 1. NHANES is a biennial national cross-sectional study conducted by the Center for Disease Control and Prevention (CDC) and the National Center for Health Statistics (NCHS), that surveys health, nutrition, and other lifestyle factors across the noninstitutionalized civilian population of the United States [16]. The study employs a multistage probability selection design to generate a nationally representative sample of the American population and to ascertain the prevalence of diseases, health outcomes, and associated environmental and behavioral risk factors [17]. Consenting participants fulfill a household screener and a home interview. The latter consists of a series of questionnaires administered in their homes that cover a range of areas, including demographic, occupational, health, and dietary-related matters. Some subjects are selected for a medical examination, which includes a variety of physical measurements, a dental examination, and biological specimens for laboratory testing. Examination data were collected in the Mobile Examination Center. In addition, dietary data were collected via 24-hour recalls to ascertain the frequency of consumption and estimate nutrient intake. Cancer, diabetes, cardiovascular disease, and renal disease statuses are assessed as self-reported items in the Medical Conditions Questionnaire (MCQ). Tumor stage data for cancer survivors are not part of the survey. All study procedures and protocols were approved by the NCHS Ethics Review Board and all participants provided informed and written consent [17].

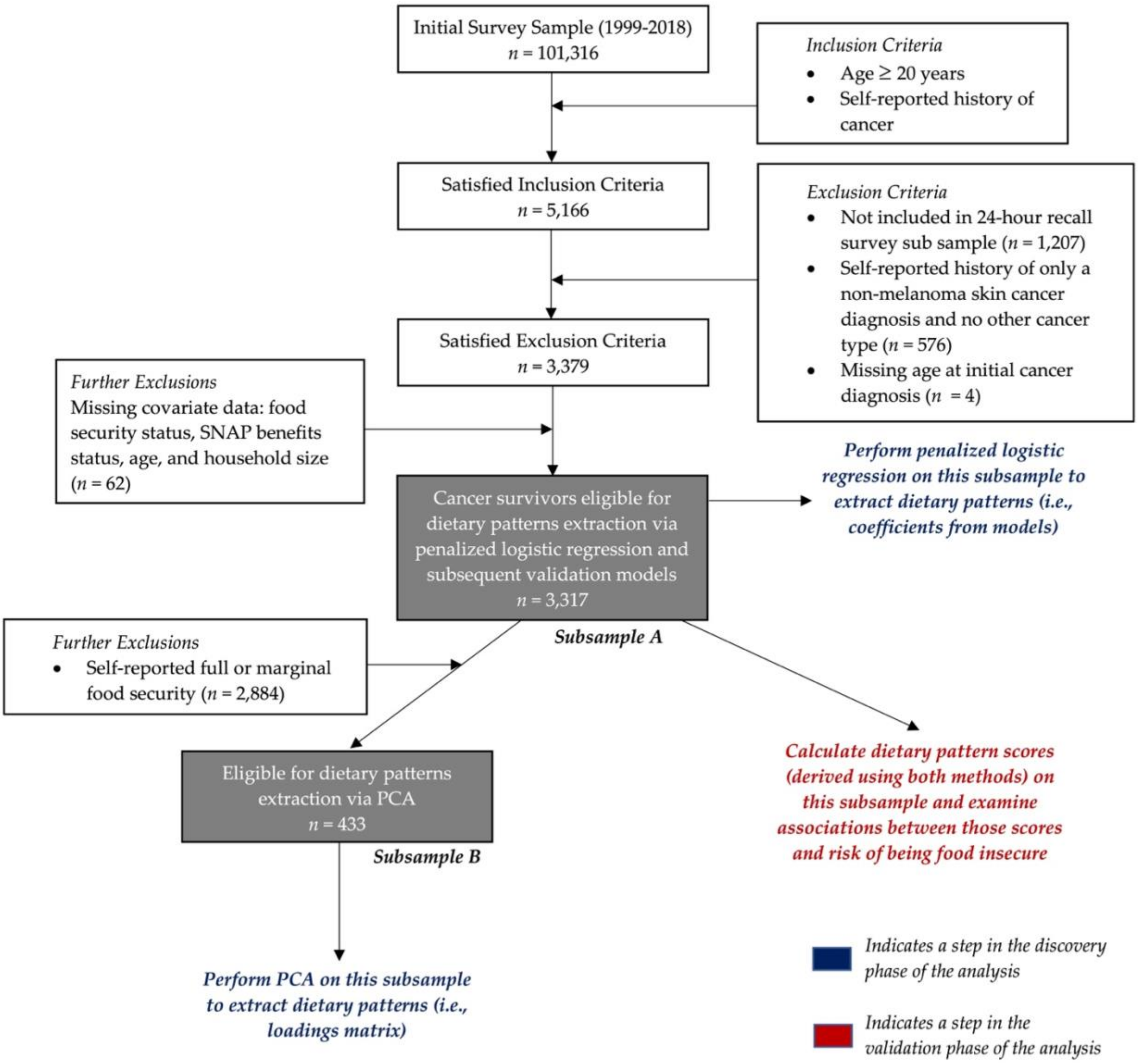

Figure 1. Study sample flow chart detailing the sample selection process and the analytical strategy. Subsamples A and B are periodically referred to in the text.

2.1. Study Sample

Figure 1 details a flow diagram of the sample selection process. We used data from ten survey cycles between 1999-2018 that included a subsample of 5,166 participants aged at least 20 years with a self-reported history of cancer and reliable dietary data, as defined by the NCHS. Cancer status and history were ascertained in the MCQ by asking "Have you ever been told by a doctor or health professional that you had cancer or a malignancy of any kind?". Individuals reporting a history of non-melanoma skin cancer (n = 576) and no other cancer type were recoded as not having a significant cancer history given the generally benign course associated with these cancers that might otherwise bias the sample [18]. Dietary patterns extraction procedures using penalized logistic regression models were performed on individuals reporting a history of a cancer diagnosis and who demonstrated complete records for food security status, data on receipt of Supplemental Nutrition Assistance Program (SNAP) benefits, household size, and age (subsample A, n = 3,317). To extract dietary patterns that characterized intake in the population of food insecure cancer survivors using PCA we further excluded individuals reporting full or marginal food security (n = 2,884) (i.e., only food insecure cancer survivors—subsample B, n = 433). Validation analyses examining the relationship between computed diet pattern indices and the risk of FI were performed on the pooled subsample of food secure and food insecure cancer survivors (subsample A, n = 3,317)

2.2. Demographic and Physical Health Covariates

Demographic characteristics were self-reported and captured in the home interview. Age was modeled continuously, and sex was coded dichotomously (Male and Female). Race and ethnicity were categorized as Mexican American, Other Hispanic, Non-Hispanic White, Non-Hispanic Black, and Other/Multiracial (although we note that our final analytical models implemented a binary-coded version given the small sample size—non-Hispanic White and non-White). We considered income status using the family income-to-poverty ratio (FIPR), classified into two categories: < 1.3 or ≥ 1.3. This value was chosen deliberately as it is a threshold commonly employed by various federal safety net programs to evaluate low socioeconomic status for program eligibility [19]. We modeled household size numerically.

Health-related and behavioral characteristics included body mass index (BMI) (modeled continuously in units of kg/m^2), smoking status, which was categorized as 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), and drinking status, which classified participants as heavy drinkers (≥ 14 g/d for women and ≥ 28 g/d for men), moderate drinkers (0.1-13.9 g/d for women and 0.1-27.9 g/d for men), and abstainers (< 0.1 g/d)—note: 1 serving of alcohol equates to 14 g of alcohol [20,21]. Finally, we computed a modified version (given limitations with the data provided through NHANES and those required for full computation of the metric) of the Charlson Comorbidity Index and weekly metabolic equivalents (MET), as previously described, to evaluate comorbidity burden and physical activity (measured as all physical activity exerted on a weekly basis), respectively [22,23]. In subsequent modeling efforts, these measures were modeled as continuous variables.

2.3. Dietary Assessment Data

Dietary data were collected using the 24-hour recall method from NHANES participants during an in-person interview (performed in the MEC) [24]. A subsequent, unannounced 24-hour recall is collected via telephone within 3-10 days following the interview. Dietary interview protocols and the administered 24-hour recall were designed to provide detailed dietary data by capturing the foods and beverages consumed by participants within the preceding 24 hours. The methodology for the dietary interview component was developed by the USDA’s Food Surveys Research Group and incorporated the USDA’s automated multiple-pass method [24,25]. Dietary data collected between 1999 and 2002 included only one day of intake from participants, whereas data collected between 2003 and 2018 included two days of recalls from each participant. To make full use of the available data and minimize any bias introduced by using a single day of dietary intake values, we averaged intake values across both days of data collection [26,27]. Daily total energy and nutrient intake data were obtained for each participant. Total energy and nutrient intake values were estimated from foods noted in the dietary interview while cross-referencing the Food and Nutrient Database for Dietary Studies [28].

Intake according to food groups data were obtained from the publicly available USDA Food Patterns Equivalents Database (FPED) and MyPyramid Equivalents Database (MPED) [29,30]. The FPED and MPED use a database of 8,356 commonly consumed food items to compute intake equivalents across 37 food pattern components. Considering this classification scheme, a modified yet similar, food-grouping scheme involving 26 food groups was adopted for this analysis. These 26 groups and the way they were collapsed are detailed in Table Supplemental Table 1. Prior to any dietary patterns extraction procedures, food group intake equivalents were divided by a subject’s total caloric intake so that a multivariate density model could be implemented to adjust for total energy intake and minimize the likelihood of confounding by total energy intake in any of the subsequent modeling efforts [31].

2.4. Cancer Status and Food Security Data

The MCQ provides survey participants with an avenue for self-reporting data on medical conditions. Time since cancer diagnosis was computed as the time elapsed between a subject's age at their first cancer diagnosis and their current age and was subsequently categorized ($<$ 2 years, 2 $\geq$ and $<$ 6 years, and $\geq$ 6 years). Participants with a history of a cancer diagnosis were grouped into their primary cancer type. That is the cancer type with the longest associated time since diagnosis. Lastly, the 32 cancer types listed in the NHANES MCQ were collapsed into a set of 8 primary cancer groups proposed by colleagues (Breast, Gastrointestinal, Genitourinary, Gynecological, Male Reproductive, Melanoma, Skin-Unknown, and Other) [32]. Again, individuals reporting only a diagnosis of non-melanoma skin cancer and no other cancer were omitted as positive cancer cases, given the generally benign course associated with this malignancy [18].

Food security status was assessed using the U.S. Food Security Survey Module (U.S. FSSM), an 18-item screener employed by NHANES since the 1999 cycle to assess food security experienced by subjects over the preceding year [33,34]. The questionnaire was administered in the home interview setting, with one adult responding on behalf of all individuals in that household, regardless of whether they were included in the survey. The survey is comprised of 10 items dedicated to households without children and eight items for households with children. Counts and affirmative responses on the questionnaire are used to bin subjects into overall food security categorizations. Those responding in the affirmative to $\leq$ 2 items were categorized as food secure, while those responding in the affirmative to $\geq$ 3 items were categorized as food insecure, and followed validated cutoffs [33,34]. Additionally, receipt of food assistance and specifically participation in the Supplemental Nutrition Assistance Program (SNAP) is reported in the U.S. FSSM. These data were captured by prompting participants on whether any household member was authorized to receive SNAP benefits in the 12 months preceding the interview.

2.5. Statistical Analysis

Descriptive statistics were tabulated on demographic variables across levels of food security/cancer status using subsample A (Figure 1). Dietary patterns were extracted using penalized logistic regression models and PCA (see Appendix A for a detailed description of the implementation of these procedures). For the former, we used four binary outcomes: food insecurity status (food insecure vs food secure), age $\geq$ 60 years, household receipt of SNAP benefits in the last 12 months, and household size $\geq$ 5, which are all associated or understood risk factors for FI [35,36]. A Pearson correlation matrix was generated to evaluate relationships between the dietary patterns and food groups in subsample A (n = 3,117). To validate the extracted dietary patterns, we used the loadings and coefficients (from the PCA and elastic net procedures, respectively) to compute dietary pattern scores for subjects identifying with a history of cancer (subsample A, n = 3,317). The validation phase of the analysis comprised the analytical goal of determining the relationship between the extracted pattern scores and the risk of FI in the cancer survivor population (Figure 1). To this end, we implemented weighted logistic regression models that modeled the log odds of being food insecure as a function of the dietary patterns scores and relevant covariates. This step included all subjects with a reported history of cancer (subsample A, n = 3,317). Alcohol consumption was not included as a covariate in these models to minimize collinearity, given that the extracted patterns already considered alcohol consumption in their computation.

We modeled the diet scores using multiple approaches to evaluate the robustness of the results. First, we modeled the scores categorically after binning participants into quintiles. A test for linear trend across the quintiles was performed by generating a new variable that assigned each subject the median value of their respective quintile and modeling it as a continuous variable. Second, we standardized the diet scores by dividing them by their respective standard deviation and then modeling them as continuous variables. Third, we added a quadratic term to the previous model to assess for divergence from a linear relationship. Lastly, we modeled the diet scores using restricted cubic splines with five knots to model the relationship flexibly and evaluate for dose-response and, again, linearity. All models were adjusted for relevant confounders including age, sex, race/ethnicity, family income-to-poverty ratio, highest level of education attained, household size, SNAP participation status, BMI, estimated caloric intake, weekly MET minutes, primary cancer site, smoking status, and the Charlson Comorbidity Index score. We fit stratified models according to sex, time since primary cancer diagnosis, and attained level of education. We accounted for the complex and multi-stage probability design of the study by following NCHS analytical guidelines and weighting our analyses accordingly [37]. All analyses were conducted at $\alpha = 0.05$ and were performed in R version 4.2.1. All accompanying R code and data files necessary to reproduce these analyses can be found at: https://github.com/cmainov/NHANES-Diet-Penalized-Regression.

3. Results

3.1. Descriptive Statistics

Sociodemographic, clinical, and behavioral characteristics are summarized in Table 1. On average, those with a reported history of cancer and with self-identified low food security were younger than food secure cancer survivors, were more likely to identify as female, live in a home with ≥ 5 individuals, and belong to a minority group compared to those identifying as food secure with a history of cancer who were older, had a more balanced ratio of the sexes, and were predominantly white. Food insecure survivors also had lower attained educational status compared to food secure individuals, tended to have a lower FIPR, reported being more physically active throughout the week, and were more likely to be receiving food assistance through SNAP in addition to, on average, consuming over 200 fewer daily calories compared to their food secure counterparts. Regarding cancer site, food insecure survivors disproportionally reported gynecological cancers as their primary form of cancer compared to a lower rate in the food secure survivors. Concomitantly, a larger proportion of male reproductive cancers were represented in the food secure sample relative to the food insecure sample. There was no gross difference observed in time-since-diagnosis across the two groups. However, food insecure individuals had a slightly larger mean Charlson Comorbidity Index score than food secure participants. Finally, those identifying as food insecure were more likely to report being current smokers than food secure individuals with cancer.

Table 1. Sociodemographic and behavioral characteristics of the cancer survivor study sample (subsample A in Figure 1), stratified by food security status. Frequencies are presented with percentages in parentheses.

Characteristic	Total Survivors (n = 3,317)	Food Insecure (n = 433)	Food Secure (n = 2,884)	p
Age				< 0.01
Mean (SD)	62.6 (14.8)	50.8 (15.7)	64.1 (14)	
Sex				< 0.01
Male	1527 (40.9)	143 (24.5)	1384 (42.9)	
Female	1790 (59.1)	290 (75.5)	1500 (57.1)	
Race/Ethnicity				< 0.01
Mexican American	235 (2.7)	67 (9.2)	168 (1.9)	

Other Hispanic	183 (2.6)	48 (5.8)	135 (2.2)	
Non-Hispanic White	2219 (84.3)	208 (68.3)	2011 (86.3)	
Non-Hispanic Black	534 (6.9)	88 (11.2)	446 (6.4)	
Other/Multiracial	146 (3.5)	22 (5.5)	124 (3.2)	
Education Attained				< 0.01
≤ High School	1577 (36.8)	279 (59.3)	1298 (34.1)	
≥ Some College	1737 (63.2)	152 (40.7)	1585 (65.9)	
FIPR				< 0.01
≥ 1.3	2279 (82.2)	128 (39.1)	2151 (87.6)	
< 1.3	800 (17.8)	288 (60.9)	512 (12.4)	
Household Size				< 0.01
< 5 Persons	3027 (92.3)	345 (79.3)	2682 (93.9)	
≥ 5 Persons	290 (7.7)	88 (20.7)	202 (6.1)	
BMI (kg/m^2)				0.23
Mean (SD)	29.2 (6.6)	29.7 (7.2)	29.1 (6.5)	
Weekly MET Minutes				<0.01
Mean (SD)	2314.2 (4475.2)	4641.1 (7771)	2034.9 (3804.1)	
Daily Caloric Intake (kcal)				<0.01
Mean (SD)	1894.6 (687)	1711.1 (740.2)	1917 (677.0)	
Charlson Comorbidity Score				< 0.01
Mean (SD)	3.0 (1.4)	3.3 (1.8)	3.0 (1.4)	
SNAP Assistance				< 0.01
No	2839 (88.6)	220 (49.9)	2619 (93.3)	
Yes	478 (11.4)	213 (50.1)	265 (6.7)	
Cancer Site				< 0.01
Breast	563 (17.2)	58 (10.2)	505 (18.0)	
Gastrointestinal	321 (7.7)	45 (11.3)	276 (7.3)	
Genitourinary	145 (3.7)	15 (3.7)	130 (3.7)	
Gynecological	522 (17.8)	132 (38.1)	390 (15.3)	
Male Reproductive	620 (13.8)	50 (6.0)	570 (14.7)	
Melanoma	240 (9.3)	15 (2.0)	225 (10.2)	
Other	592 (19.1)	99 (23.1)	493 (18.6)	
Years Since Diagnosis				0.53
< 2 years	817 (22.0)	113 (21.4)	704 (22)	
≥ 2 and < 6 years	1991 (64.4)	257 (67.4)	1734 (64.1)	
≥ 6 years	497 (13.6)	60 (11.2)	437 (13.9)	
Smoking Status				< 0.01
Current	517 (16.4)	142 (37.8)	375 (13.8)	
Former	1347 (38.9)	120 (26.8)	1227 (40.4)	
Never	1451 (44.7)	170 (35.4)	1281 (45.8)	
Alcohol Use				0.13
Heavy	323 (12.8)	29 (6.6)	294 (13.6)	

Moderate	498 (16.1)	48 (15.0)	450 (16.2)
None-drinking	2496 (71.1)	356 (78.5)	2140 (70.2)

Percentages may not add to 100% given rounding

p values are from chi-square tests for categorical variables and t-tests for continuous variables

3.2. Discovery Phase: Dietary Patterns Extraction

There were six dietary patterns extracted from both procedures. The patterns derived using penalized logistic regression were named according to the outcome variable used in each model (we named these the Food Insecurity (FI), Age, SNAP, and Household Size patterns, respectively). Figure 2 illustrates the optimal combinations of the model tuning parameters, λ and α, that were ultimately selected for each model. For the model with FI as the response variable, the LASSO regression ($\alpha = 1$) solution was optimal while the ridge regression solution ($\alpha = 0$) was optimal for the model with household size as the response. The models with age and SNAP benefits as the outcomes yielded optimized solutions with α in the elastic net range, $\alpha \in (0,1)$. The coefficients for each of these models are found in Supplementary Table 2. We note that the coefficients for several food groups shrunk to zero, effectively eliminating them from subsequent score computations.

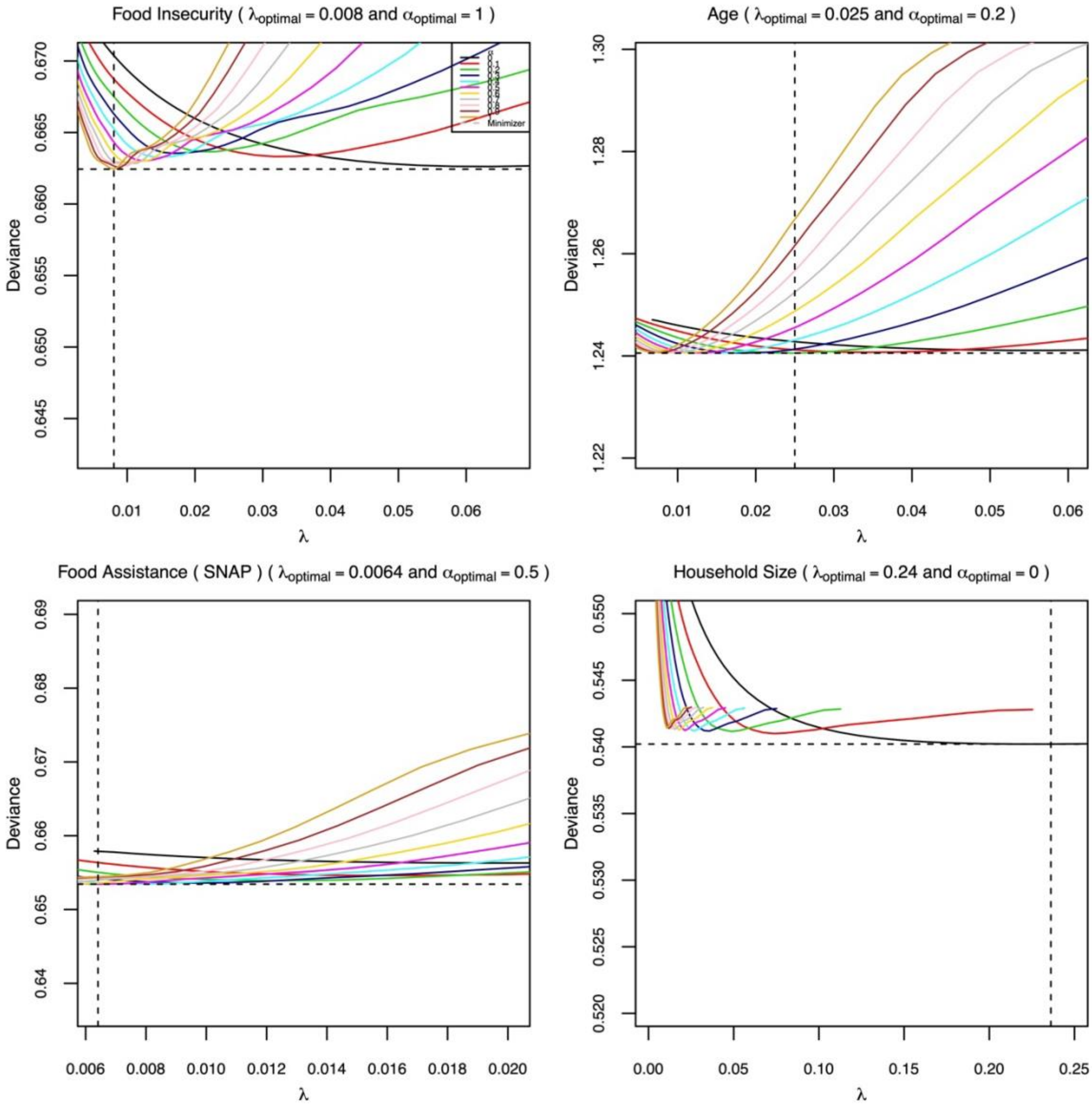

Figure 2. Optimal Combinations of α and λ (minimizers) in the penalized logistic regression models used for dietary patterns extraction performed on subsample A *(n* = 3,117).

In Table 2, we detail the Pearson correlation coefficients amongst pattern scores and food groups. We used a cut-off threshold of |0.30| to identify food groups that significantly contributed to these patterns [38]. The FI pattern was positively correlated with intakes of potatoes, and added sugars while moderately and negatively correlated with intake of other vegetables. The Age pattern was positively correlated with intakes of milk, fruit, and whole grains while negatively correlated with cheese. Overall, this pattern was negatively correlated with the FI pattern (r = -0.28). The SNAP dietary pattern was strongly and positively correlated with the FI pattern (r = 0.80) as well as with added sugars while being negatively correlated with alcohol, dark-yellow vegetables, other vegetables, and nuts. The final Household Size pattern was also strongly and positively correlated with the FI pattern (r = 0.63) and negatively correlated with intakes of yogurt, other fruit, citrus, melons and berries, tomatoes, and other vegetables being positively correlated with intake of added sugars.

Table 2. Pearson correlation coefficients matrix amongst each of the derived patterns (extracted using either penalized logistic regression or principal components analysis) and the food groups used in the analysis. A lower triangular

Pearson correlation matrix is appended, showing the correlations between the extracted dietary patterns. This analysis was performed on subsample A (Figure 1), an analytical subsample of cancer survivors (*n* = 3,317).

Pattern / Food Groups	Food Insecurity (FI)†	Age†	Food Assistance (SNAP)†	Household Size†	Modified Western‡	Prudent‡
Processed Meats	-0.05	-0.01	0.04	0.03	0.12	-0.22
Meats	0.22	-0.03	0.08	0.00	0.07	-0.17
Poultry	0.00	-0.26	-0.08	0.20	-0.03	**0.35**
Seafood—High n-3	-0.16	0.05	-0.11	-0.06	-0.04	**0.30**
Seafood—Low n-3	-0.17	0.08	-0.06	-0.16	-0.04	0.07
Eggs	0.07	0.11	0.00	0.15	0.24	0.07
Solid Fats	0.12	0.04	0.22	-0.12	0.21	-0.46
Oils	-0.26	0.04	-0.24	0.09	**0.34**	0.12
Milk	-0.10	**0.37**	0.00	-0.19	-0.06	0.12
Yogurt	-0.09	0.05	-0.10	**-0.32**	0.07	**0.35**
Cheese	-0.06	**-0.39**	0.05	-0.19	**0.34**	-0.28
Alcohol	-0.19	-0.27	**-0.34**	-0.09	**-0.36**	-0.16
Fruit—Other	-0.23	**0.41**	-0.24	**-0.33**	0.03	**0.49**
Fruit—Citrus, melons, and berries	-0.20	0.18	-0.19	**-0.36**	0.04	**0.50**
Tomatoes	-0.21	0.04	-0.17	**-0.36**	**0.48**	0.14
Dark-Green Vegetables	-0.21	-0.19	-0.26	-0.22	0.26	**0.53**
Dark-Yellow Vegetables	-0.16	0.10	**-0.34**	-0.06	0.14	**0.45**
Other Vegetables	**-0.50**	0.17	**-0.65**	**-0.48**	**0.46**	**0.45**
Potatoes	**0.41**	0.25	0.06	0.05	0.16	-0.04
Other Starchy Vegetables	-0.03	0.16	-0.11	-0.15	-0.12	0.17
Legumes	0.01	-0.24	0.21	0.23	0.04	-0.08
Soy	-0.08	-0.11	-0.20	0.22	0.08	0.21
Refined Grains	-0.13	-0.12	0.17	0.13	0.11	**-0.34**
Whole Grains	-0.20	**0.47**	-0.25	-0.27	-0.05	**0.38**
Nuts	-0.28	0.10	**-0.31**	-0.02	0.18	0.19
Added Sugars	**0.76**	-0.28	**0.64**	**0.48**	**-0.32**	-0.27
FI	--					
Age	-0.28	--				
SNAP	0.80	-0.37	--			
Household Size	0.63	-0.50	0.62	--		
Modified Western	-0.26	0.09	-0.29	-0.31	--	
Prudent	-0.40	0.35	-0.56	-0.41	0.16	--

† Dietary pattern obtained using penalized logistic regression.

‡ Dietary pattern obtained using principal components analysis (PCA).

Correlation coefficients (*r*) ≥ |0.30| are bolded to ease the identification of notable food groups characterizing the different patterns.

For the patterns extracted with PCA, we evaluated a scree plot initially and found that an "elbow" appeared after the fourth principal component (Supplementary Figure 1). However, upon evaluation of the component loading matrix (Supplementary Table 2) and the table of correlations (Table 2), only the first and second principal components had interpretable loadings that were deemed meaningful. Thus, a decision was made to retain only the first two components given the weight placed on having interpretable components [39]. The eigenvalues suggested that these first two components accounted for 14.1% of the variation in the 24-hour recall data. Both patterns shared similarities in that both were positively correlated with vegetable consumption and negatively correlated with added sugar and alcohol. However, while the first principal component emphasized modest meat, processed meat, solid fat, oil, eggs, milk, cheese, potato, soy, nuts, and refined grains consumption, the second principal component emphasized fruit, poultry, eggs, milk, yogurt, high n-3 seafood, soy, and whole grains. The second principal component was also negatively correlated with meat and processed meat consumption, cheese, solid fat, cheese, legumes, and refined grains intakes and, overall, had stronger positive correlations to whole grains, fruit, soy, nuts, and vegetable intake compared to the first component. Given both healthful and unhealthful aspects of the first principal component, we termed this pattern the Modified Western pattern [40,41]. In contrast, the second principal component was termed the Prudent pattern, given its greater and more consistent emphasis on the pillars of healthful eating cited previously in the literature [42].

Differences across sociodemographic covariates between high and low median splits of each of the six dietary patterns in the subsample of cancer survivors (subsample A, n = 3,317) are presented in Table 3. On average, those with higher scores on the Age pattern tended to be older. Subjects with greater scores on the FI, SNAP, and Household Size patterns also tended to be younger and have a lower FIPR than those with lower pattern scores. Subjects with high scores on the household size pattern were also more likely to report living in a home with ≥ 5 persons compared to low scorers, while high scorers on the FI and SNAP patterns were more likely to identify as food insecure and receive SNAP benefits compared to low scorers. Finally, high scorers on the Prudent pattern were, on average, more likely to report as never smokers and less likely to report as current smokers compared to low scorers.

Table 3. Derived dietary patterns and their distributions across study covariates in cancer survivors (subsample A in Figure 1) included in the survival analysis (n = 3,317).

	Food Insecurity (FI) Pattern†		**Age Pattern**†		**Food Assistance (SNAP) Pattern**†		**Household Size Pattern**†		**Modified Western**‡		**Prudent**‡	
	M1 (n = 1658)	**M2** (n = 1659)	**M1** (n = 1658)	**M2** (n = 1659)	**M1** (n = 1658)	**M2** (n = 1659)	**M1** (n = 1658)	**M2** (n = 1659)	**M1** (n = 1658)	**M2** (n = 1659)	**M1** (n = 1658)	**M2** (n = 1659)
Age		**		**		**		**				**
Mean (SD)	64.1 (13.5)	61.1 (15.9)	58.6 (14.9)	67.4 (13.1)	64.5 (13.3)	60.5 (16.0)	65.2 (13.3)	59.8 (15.8)	63 (15.1)	62.3 (14.5)	60.3 (15.4)	65.1 (13.6)
Sex										*		**
Male	752 (41.0)	775 (40.8)	741 (41.1)	786 (40.6)	762 (41.1)	765 (40.6)	741 (38.9)	786 (43.0)	821 (44.1)	706 (38.2)	836 (46.9)	691 (34.4)
Female	906 (59.0)	884 (59.2)	917 (58.9)	873 (59.4)	896 (58.9)	894 (59.4)	917 (61.1)	873 (57.0)	837 (55.9)	953 (61.8)	822 (53.1)	968 (65.6)
Race/Ethnicity		*		**		**		**		**		*
Minority	513 (14.2)	585 (17.2)	663 (18.2)	435 (12.6)	474 (12.8)	624 (18.8)	428 (11.8)	670 (19.8)	639 (19.3)	459 (12.6)	528 (14.2)	570 (17.3)
White	1145 (85.8)	1074 (82.8)	995 (81.8)	1224 (87.4)	1184 (87.2)	1035 (81.2)	1230 (88.2)	989 (80.2)	1019 (80.7)	1200 (87.4)	1130 (85.8)	1089 (82.7)
Education		**		*		**		**				**
≤ HS	691 (31.5)	886 (42.6)	752 (34.4)	825 (39.6)	700 (30.3)	877 (43.9)	722 (33.3)	855 (40.6)	800 (37.6)	777 (36.2)	867 (40.2)	710 (33.2)
≥ Some College	966 (68.5)	771 (57.4)	905 (65.6)	832 (60.4)	957 (69.7)	780 (56.1)	934 (66.7)	803 (59.4)	856 (62.4)	881 (63.8)	791 (59.8)	946 (66.8)
FIPR		**				**		**		**		*
≥ 1.3	1214 (86.9)	1065 (77.2)	1104 (81.4)	1175 (83.2)	1232 (87.5)	1047 (76.4)	1189 (84.9)	1090 (79.4)	1100 (79.2)	1179 (84.8)	1089 (80.2)	1190 (84.5)
< 1.3	327 (13.1)	473 (22.8)	443 (18.6)	357 (16.8)	307 (12.5)	493 (23.6)	342 (15.1)	458 (20.6)	437 (20.8)	363 (15.2)	462 (19.8)	338 (15.5)
Household size												

< 5 Persons	1547 (94.5)	1480 (90.0)	1460 (90.0)	1567 (95.1)	1563 (95.4)	1464 (89.0)	1557 (96.6)	1470 (87.7)	1495 (91.8)	1532 (92.8)	1482 (90.4)	1545 (94.4)
≥ 5 Persons	111 (5.5)	179 (10.0)	198 (10.0)	92 (4.9)	95 (4.6)	195 (11.0)	101 (3.4)	189 (12.3)	163 (8.2)	127 (7.2)	176 (9.6)	114 (5.6)
BMI												
Mean (SD)	29.3 (6.6)	29 (6.6)	29.3 (6.8)	29 (6.4)	29 (6.3)	29.4 (6.9)	29.2 (6.3)	29.1 (6.9)	28.6 (6.2)	29.7 (6.9)	29.7 (6.9)	28.6 (6.2)
Weekly MET Minutes				*								
Mean (SD)	2185.6 (3865.4)	2454.4 (5054.7)	2611.5 (4911.5)	1959 (3862.2)	2117.8 (3584.0)	2529.2 (5274.0)	2056.1 (3644.8)	2593.5 (5214.2)	2313.3 (4364.1)	2314.9 (4567.9)	2504.3 (5077.2)	2108.2 (3705.2)
Daily Caloric Intake												
Mean (SD)	1836.2 (660.8)	1958.3 (709.2)	1938.4 (696.4)	1842.8 (672.3)	1842.1 (655.0)	1952.2 (716.2)	1848 (655.0)	1945.1 (716.8)	1898.6 (705.1)	1891.3 (671.6)	2041.2 (746.9)	1736.6 (575.7)
CCI												
Mean (SD)	3.1 (1.4)	3.0 (1.4)	2.9 (1.3)	3.1 (1.5)	3.0 (1.4)	3.0 (1.4)	3.0 (1.4)	3.0 (1.4)	3.0 (1.4)	3.0 (1.4)	3.0 (1.4)	3.0 (1.4)
Food Security												
Food Secure	1482 (93.7)	1357 (83.1)	1351 (85.4)	1488 (92.5)	1512 (95.0)	1327 (81.7)	1476 (92.9)	1363 (84.0)	1379 (86.5)	1460 (90.5)	1352 (84.7)	1487 (92.9)
Food Insecure	176 (6.3)	302 (16.9)	307 (14.6)	171 (7.5)	146 (5.0)	332 (18.3)	182 (7.1)	296 (16.0)	279 (13.5)	199 (9.5)	306 (15.3)	172 (7.1)
SNAP Assistance				*								
No	1482 (93.7)	1357 (83.1)	1351 (85.4)	1488 (92.5)	1512 (95.0)	1327 (81.7)	1476 (92.9)	1363 (84.0)	1379 (86.5)	1460 (90.5)	1352 (84.7)	1487 (92.9)
Yes	176 (6.3)	302 (16.9)	307 (14.6)	171 (7.5)	146 (5.0)	332 (18.3)	182 (7.1)	296 (16.0)	279 (13.5)	199 (9.5)	306 (15.3)	172 (7.1)
Smoking Status						*						**
Current	188 (11.4)	329 (21.9)	344 (20.6)	173 (11.4)	191 (11.4)	326 (21.9)	179 (11.3)	338 (21.9)	284 (18.3)	233 (14.8)	393 (24.5)	124 (7.6)
Former	724 (42.3)	623 (35.2)	628 (36.3)	719 (42.0)	712 (42.6)	635 (34.8)	727 (43.0)	620 (34.4)	664 (38.6)	683 (39.2)	670 (37.8)	677 (40.1)

Never	745 (46.3)	706 (43.0)	685 (43.1)	766 (46.6)	754 (46.0)	697 (43.3)	750 (45.7)	701 (43.6)	709 (43.2)	742 (46.0)	594 (37.7)	857 (52.2)
Alcohol Use				*								
Heavy	214 (16.3)	109 (9.1)	242 (18.9)	81 (5.6)	265 (20.7)	58 (4.2)	206 (16.1)	117 (9.3)	227 (19.9)	96 (6.9)	207 (15.2)	116 (10.2)
Moderate	272 (16.3)	226 (15.9)	264 (17.2)	234 (14.8)	294 (17.3)	204 (14.8)	266 (15.0)	232 (17.3)	276 (17.4)	222 (15.0)	250 (16.9)	248 (15.3)
Non-drinking	1172 (67.5)	1324 (75.0)	1152 (63.9)	1344 (79.6)	1099 (62.0)	1397 (81.0)	1186 (68.9)	1310 (73.5)	1155 (62.7)	1341 (78.1)	1201 (67.9)	1295 (74.5)

M1 refers to the lower 50% fraction of the data while M2 refers to the upper 50% fraction of the data after splitting the diet scores at the median.

† Dietary pattern obtained using penalized logistic regression.

‡ Dietary pattern obtained using principal components analysis (PCA).

3.3. Validation Phase: Logistic Regression

Using binary logistic regression models, we found, after multivariable adjustment, significant associations between the extracted pattern scores and the odds of being food insecure (Table 4). The FI, SNAP, and Household Size patterns were all strongly and positively associated with the risk of being food insecure. Among those, the FI pattern had the most considerable magnitude of association, with the odds of FI being 2.42-fold greater in the fifth quintile compared to the first quintile. Similarly, all three patterns had similar magnitudes of association when the diet score was modeled linearly. For the FI pattern, a one standard deviation increase in the score was associated with 50% increase in the odds of being food insecure. Concerning associations in the opposite direction, only the Prudent pattern was inversely associated with FI, with the highest quintile observing a 60% reduction in the odds of being food insecure compared to the first quintile. A one standard deviation increase in this pattern scores was also significantly associated with a 24% decrease in the odds of being food insecure. For all the noted dietary patterns, tests for linear trends revealed linear behavior, in their respective directions, across the quintiles, and these findings were generally supported by the results from fitting models with restricted cubic splines (Figure 3), although the strongest relationship, again, appeared to belong to the FI pattern.

Table 4. Odds ratios[†] and 95% confidence intervals for the relationship between the dietary patterns scores and the odds of being food insecure. There were 3,317 cancer survivors (subsample A in Figure 1) that contributed to this analysis.

Dietary Pattern[a]	Q1	Q2	Q3	Q4	Q5	$p_{Q5\text{-}Q1}$	p_{trend}	$HR^{b}_{continuous}$	$p^{c}_{quadratic}$
Food Insecurity[†]	1.00	1.09 (0.58-2.02)	1.18 (0.57-2.45)	1.91 (1.04-3.53)*	2.42 (1.21-4.82)*	0.01*	<0.01**	1.50 (1.19-1.90)**	0.11
Age[†]	1.00	1.91 (1.12-3.27)*	1.41 (0.67-2.93)	1.14 (0.49-2.69)	1.82 (0.93-3.56)	0.08	0.28	1.05 (0.87-1.27)	0.57
Food Assistance (SNAP)[†]	1.00	1.38 (0.65-2.93)	1.44 (0.77-2.71)	2.54 (1.22-5.30)*	2.23 (1.26-3.94)**	<0.01**	<0.01**	1.37 (1.12-1.68)**	0.46
Household Size[†]	1.00	1.63 (0.78-3.43)	1.00 (0.52-1.92)	2.77 (1.46-5.25)**	2.02 (0.98-4.18)	0.06	0.01*	1.27 (1.04-1.54)*	0.36
Modified Western[‡]	1.00	0.86 (0.48-1.51)	0.69 (0.33-1.45)	1.46 (0.81-2.64)	1.33 (0.66-2.67)	0.42	0.16	1.05 (0.88-1.25)	0.70
Prudent[‡]	1.00	0.81 (0.37-1.78)	1.09 (0.53-2.26)	0.54 (0.26-1.10)	0.40 (0.20-0.80)**	<0.01**	<0.01**	0.76 (0.63-0.92)**	0.27

** $p < 0.01$

* $p < 0.05$

[a] All models adjusted for age, sex, race/ethnicity, family income-to-poverty ratio, highest level of education attained, household size, SNAP participation status, BMI, estimated caloric intake, weekly MET minutes, primary cancer site, smoking status, and the Charlson Comorbidity Index score and were weighted according to guidelines provided by the NCHS.

[b] Hazard ratio (HR) corresponding to a standard deviation increase in the diet pattern score.

[c] Wald test *p*-value for a quadratic polynomial term.

[†] Dietary pattern obtained using penalized logistic regression.

[‡] Dietary pattern obtained using principal components analysis (PCA).

349

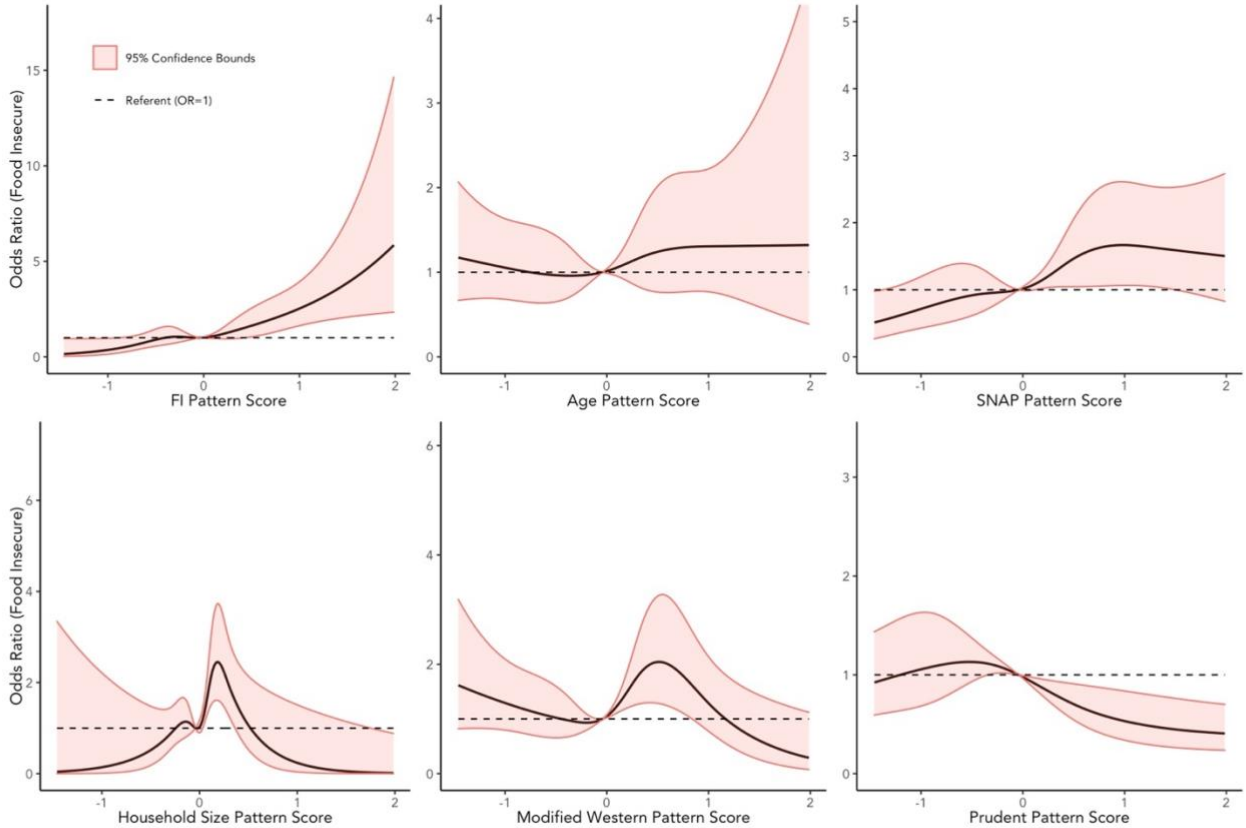

Figure 3. Adjusted restricted cubic spline curves demonstrating the relationships between dietary pattern scores and the odds of being food insecure in the subsample of cancer survivors (subsample A in Figure 1). These models used five knots to model each dietary pattern score and adjusted for age, sex, race/ethnicity, income to poverty ratio, highest level of education attained, household size, SNAP participation status, BMI, estimated caloric intake, weekly MET minutes, smoking status, and the Charlson Comorbidity Index (CCI) score and were weighted using normalized weights. The hazard at the median of each dietary pattern score was employed as the referent.

Results from stratified models are presented in Supplementary Table 3. We found that relationships between each diet pattern score and FI risk were significantly stronger in females than males. The highest quintile of the FI pattern demonstrated a 3.48-fold greater risk of being food insecure compared to the lowest quintile, while the signal in the male population was blunted with only a non-significant 1.46-fold greater risk of FI in the fifth relative to the lowest quintile. Likewise, comparisons between the SNAP, Household Size, and Prudent patterns displayed similar phenomena (Supplementary Table 3). When comparing time since primary cancer diagnosis, we found that the FI pattern was associated with 4.72-fold greater risk of FI in those subjects intermediately removed from a cancer diagnosis at greater than two and less than six years removed from their diagnosis. Within this group, the strongest association belonged to the SNAP pattern, where the fifth quintile demonstrated a 7.90-fold greater risk of FI than the lowest quintile, and a one standard deviation increase in pattern score was associated with a significant 1.65-fold increased risk of FI. In those ≥ 6 years removed from their primary diagnosis, there was a positive relationship between greater adherence to the FI pattern and a greater risk of FI. Moreover, the highest quintile of the prudent pattern was associated with a 67% reduction in the risk of being food insecure compared to the first quintile and there was evidence of a significant linear trend. Finally, when examining education status, it was revealed that strong and significant associations were present in the FI, SNAP, Household Size, and prudent patterns for those reporting some level of college or greater but not those with only a high school education or less.

4. Discussion

The results we present highlight major dietary patterns associated with FI in the cancer survivor population, a population plagued by high nutritional requirements and, often because of treatment-related or other side effects, limited nutritional intake. Using a combination of empirical methods, we extracted six dietary patterns to characterize the dietary intake patterns of this population. We used supervised learning in the form of penalized logistic regression to model FI and other risk factors of FI by regressing them on the 26 food groups considered in the analysis. Some of the resulting patterns were similar and consistent in that three emphasized comparable food groups, all to a similar extent, although they contained notable differences. Namely, high consumption of added sugars and low consumption of various classes of whole fruits and vegetables were themes consistently seen in those patterns (FI, SNAP, and Household Size patterns). Decreased consumption of whole grains, nuts, and legumes also highlighted these patterns, which, taken together, may suggest that food insecure survivors were, on average, more likely to be following a diet comprised, principally, of processed foods. Regarding their relationship to FI in the cancer survivor population, we found that the FI, SNAP, and Household Size patterns, in particular, were strongly and positively associated with the risk of FI.

Within the broader context of studies addressing FI and diet quality, we found that those patterns extracted from the data using penalized logistic regression shared many similarities with similar studies done in other populations. In particular, the strong relationship between the extracted patterns and intake of added sugars is consistent with reports stressing the pervasiveness of sugar-sweetened beverages and added sugar consumption in other food insecure populations [43–47]. Furthermore, though several classical studies employing unsupervised learning methods such as PCA to extract dietary patterns empirically have consistently yielded "Western-style" patterns that highlight high consumption of meat and processed meat, this was not a consistent finding in our study. Meat intake was emphasized to a modest degree in the patterns we extracted with penalized logistic regression. Nonetheless, this finding, supported by evidence elsewhere in the literature, may highlight restraint on the part of food insecure individuals from purchasing more cost-prohibitive food items, such as meat, and resorting to other low-cost and high-calorie alternatives instead [48]. All in all, we find that the clinically meaningful evidence we describe lends further support and validation of the penalized logistic regression approach as a viable alternative for extracting dietary patterns that are outcome specific.

As a comparative analysis, we implemented unsupervised learning in the form of PCA to derive dietary patterns. PCA is a powerful tool that is an established method of deriving dietary patterns but also suffers from limitations. For instance, the interpretability of the principal components may be equivocal [49]. Moreover, PCA may not always be a suitable approach for extracting patterns associated with a condition or disease outcome. This notion was substantiated by the results of our study as well. Given that the procedure only aims to constrain as much of the variation in the dietary intake data onto a single dimension, predictive potential is not a guaranteed result when implementing this approach [49]. We found that the patterns extracted using penalized regression were more consistent with previous reports in the literature detailing diet quality in other food insecure subpopulations. Moreover, considering the supporting studies we describe above, we stress that in our study, PCA did not yield dietary patterns consistent with diet quality patterns described in other food insecure populations.

In the validation phase of our analysis, we found that the FI, SNAP, and Household Size patterns were positively and strongly associated with FI, while the Prudent pattern was strongly and inversely associated with being food insecure in the cancer survivor population after controlling for several relevant confounders. Stratified analyses yielded peculiar findings. Notably, we found that stratifying the validation models by sex revealed strong effect sizes for the aforementioned association in females but not males. This finding was particularly interesting when evaluated in the context of preceding studies reporting sex-specific disparities within FI research. FI has been demonstrated to be a highly gendered and sex-specific outcome that disproportionately affects females and, specifically, females that head households as opposed to male-headed households [50]. Downstream of FI itself, it is also understood that FI impacts males and females disparately concerning clinical outcomes, with food insecure females being significantly more likely to experience obesity compared to their food insecure male counterparts [51–55]. A biological basis for explaining these disparate associations is not readily accessible, with some in the field suggesting that these relationships may be explained by the gendered societal norms concerning childcare traditionally imposed on women [51]. We posit that our results may provide an additional layer of evidence for understanding the dynamic between FI and sex-specific disparities, though we are limited in our conclusions given the potential for reverse causality due to the cross-sectional design. Nevertheless, within the context of the food insecure cancer population, this conjecture would only help to understand those relationships in younger cancer survivors of child-bearing age and not necessarily post-menopausal survivors.

In addition to strong effect sizes in females, there were disparities across time since primary diagnosis that emerged. The association between the penalized logistic regression patterns and FI was strongest in those 2-6 years removed from a cancer diagnosis. These findings suggest that dietary intake may be more relevant for predicting FI status among cancer survivors within this group and not necessarily those proximal or distal to a diagnosis. However, previous findings in the literature have not found a significant moderating effect of time since diagnosis on FI-related outcomes [8,56]. Therefore, it is challenging to clarify how time since diagnosis may be moderating the results in our validation models. Nonetheless, our results may be consistent with the hypothesis that any FI resulting from financial hardship encountered throughout the cancer care continuum may not impact survivors immediately and may persist for several years before abating. Though the results have been mixed among different studies, we believe that this is an area that requires further scrutiny if we are to understand the dynamics of food insecurity throughout the cancer care continuum.

The results we present have global public health ramifications. Clinically, FI continues to be an underappreciated social determinant of health, mainly afflicting low-income populations. A consequence of FI manifests in the trade-offs exacted on food insecure survivors when faced with competing demands of nutrition and medical care [57]. Furthermore, there are currently no known recommendations or guidelines from any influential medical association or organization stressing the need for food security screenings in this population, again underscoring the urgency and relevance of this research [57]. It was previously shown that the rates of FI in the cancer population may be substantial in the low-income cancer population compared to the general population [58]. Social and economic factors are especially crucial in prognosis and survival following diagnosis, and nutrition may be a mediating factor in survivorship. Moreover, it is imperative to underscore that compared to food secure cancer patients, food insecure cancer patients comprised a substantially larger proportion of individuals from minority racial and ethnic groups, which is also consistent with what has previously been reported [59]. Though this analysis was completed with data from the U.S., we believe that many of these findings and considerations are germane in a global context. In particular, these findings may be relevant for clinicians and cancer survivors in countries without universally subsidized health care, like the U.S. Nevertheless, many facets of cancer survivorship, such as job loss and physical disability, still define the QOL cancer survivors experience globally and are also factors that may impede access to healthy and nutritious food. Finally, this work utilizing penalized logistic regression also corroborates a novel and pre-existing framework for evaluating dietary patterns associated with particular exposures [14,15]. This approach may be helpful not only for evaluating the dietary patterns of specific populations, as we have demonstrated here but also for monitoring and evaluating the effects of nutrition policy initiatives in the U.S. and globally.

Considering the study's findings within the framework of guidelines established in the WCRF/AICR third expert report, we conclude that the cancer food insecure population within the United States may be hindered from meeting the report's benchmarks. The report stresses the vital role of fruits, vegetables, legumes, and whole grains in the prevention of incident cancer, cancer control, and bolstered survivorship [9]. We found that dietary patterns derived in both manners, although more robustly with penalized logistic regression, suggested that FI in the cancer survivor population was associated with poor dietary quality that was not aligned with those guidelines. Future studies, specifically those with longitudinal cohort designs, should more closely examine the relationships between FI and dietary intake with prognostic outcomes in this population. Nevertheless, this research elevates the importance of utilizing the WCRF/AICR guidelines in clinical settings and, in particular, subsequent to food insecurity screenings.

This analysis has several strengths, including the large, combined sample size, nationally representative sampling, control for other confounding variables, and the use of a validated module for measuring food security status. There are weaknesses in our study worth noting. As is characteristic in observational studies, residual confounding and the presence of reverse causality cannot be ruled out, particularly given the cross-sectional study design. Whether FI caused the observed dietary patterns or vice versa is not a conjecture we can explicitly arrive at with these data. Furthermore, the use of a household FI metric is essential to consider, given that FI at the household level may impart unequal burdens on its residents and, in addition, the fact that these data are self-reported in nature. With regard to dietary intake measurements, we cannot rule out any systematic biases introduced by the dietary measurement protocol. In this vain, using a 24-hour recall instead of a more robust measure of dietary intake, such as a food-frequency questionnaire, is also a notable limitation. Moreover, we must again stress that there are, to our knowledge, no current published design-based modeling software allowing users to perform penalized regression (e.g., Ridge or LASSO regression) on complex survey data. Nonetheless, as we did in our analysis, weighting those procedures with normalized weights was a deliberate strategy for curtailing any parameter or standard error bias introduced by not using all components of the complex survey design [60,61]. Concerning the use of *a posteriori* methods for dietary patterns extraction, we also concede that these methods are limited in that they do not allow us to make explicit recommendations on absolute values of dietary intake for any given food group analyzed, in contrast to some *a priori* diet quality indices. Finally, we must also consider that selection bias may arise when we include a greater proportion of individuals further removed from their diagnosis that may have less aggressive or more treatable forms of cancer that may preclude them from impaired eating, experiencing debilitating cachexia, or otherwise worse prognoses.

5. Conclusions

In summary, we conclude that dietary intake in the food insecure cancer population may be nutritionally inadequate, as measured by guidelines from numerous national institutions and organizations, and is characterized by consumption of processed and unhealthful foods with a concomitant dearth of fruits and vegetables [9,21]. These deficiencies are essential to highlight in a nutritionally vulnerable population already suspectable to malnutrition as they may lend themselves to poorer clinical outcomes, though further evidence is warranted. In addition to evaluating the effects of these dietary patterns on clinical outcomes, future studies, particularly prospective longitudinal cohort studies, are needed to highlight the impact that nutritional consequences of FI have on cancer-related outcomes. Ultimately, the results of this analysis reinforce the notion of food security as a critical social determinant of health with consequences to nutritional intake that may require persistent screenings. These findings are critical and impactful given that there are currently no best-practice guidelines or consensus criteria within the cancer survivor population to ultimately abrogate the prevalence of FI and bolster patient prognoses [57].

Supplementary Materials: The following supporting information can be downloaded at: www.mdpi.com/xxx/s1; Figure S1: Scree plot from the PCA dietary extractions procedure performed on subsample B (n = 433); Table S1: Food grouping scheme used in the present analysis.; Table S2: Model coefficients or component loadings for each of the derived patterns (extracted using either penalized logistic regression or principal components analysis) and the food groups used in the analysis; Table S3: Stratified odds ratios and 95% confidence intervals for the relationship between the dietary patterns scores and the odds of being food insecure

Author Contributions: Conceptualization, CAM.; methodology, CAM, RZ.; software, CAM, RZ.; formal analysis, CAM, RZ; investigation, CAM, FG, AEA; data curation, CAM.; writing—original draft preparation, CAM, ABO.; writing—review and editing, CAM, RZ, ABO, FG, AEA.; supervision, AEA; project administration, CAM. All authors have read and agreed to the published version of the manuscript.

Funding: CAM was supported by a research scholarship from the Health Policy Research Scholars Program at the Robert Wood Johnson Foundation.

Institutional Review Board Statement: All study procedures and protocols were approved by the NCHS Ethics Review Board and all participants provided informed consent.

Informed Consent Statement: Informed and written consent was obtained from all subjects involved in the study.

Data Availability Statement: All data used in the analyses are publicly available from the Centers for Disease Control and Prevention (https://wwwn.cdc.gov/nchs/nhanes/Default.aspx). Additionally, R code and data used specifically in these analyses are also available in the following GitHub repository: https://github.com/cmainov/NHANES-Diet-Penalized-Regression.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

A.1 Dietary Patterns Extraction Procedures: PCA

PCA, a dimension-reduction procedure commonly employed in data-driven methods for ascertaining dietary patterns in epidemiologic studies, was selected as one method for deriving dietary patterns from the 26 centered and standardized food group variables (Table 2) [62,63]. In PCA, eigen decomposition of the covariance matrix containing the predictor variables of interest yields a set of eigenvectors (containing the parameters or weights that multiply each of the n variables as a linear combination) and their corresponding eigenvalues [64]. In this sense, the first eigenvector represents a projection that maps the original data onto a new vector space (i.e., reducing it onto a single dimension) with the additional quality that it retains as much of the variance in the original data and whose eigenvalue represents its variance [64]. Geometrically, the goal of PCA can be explained as creating a set of orthogonal projections on the data that explain as much of variance in the set of predictors. Eigenvalues, a scree plot, and general interpretability of the components were used to guide decisions on the number of components to retain (see Supplementary Figure 1.) We accounted for the complex sampling design by implementing dietary patterns extraction via PCA using the *svyprcomp* function from the *survey* package in R [65]. Dietary patterns extraction using PCA was implemented on subsample B. A loadings matrix from this procedure is found appended to Supplementary Table 2.

A.2 Dietary Patterns Extraction: Penalized Logistic Regression

Regularized regression models introduce a penalty term to the likelihood function for estimating model parameters in a variety of regression frameworks [66]. Addition of a penalty effectively shrinks parameter coefficients as well as their associated variances, which is particularly useful in high-dimensional settings or in the presence of collinearity. This yields a set of more interpretable and well-behaved parameter estimates [67]. In the context of dietary patterns analysis, whereby there may be substantial collinearity amongst food groups, this becomes notably advantageous. The penalty term is added to the likelihood function, when solving the logistic regression problem, in the following form:

$$\ell(\beta) = \sum_{i=1}^{n} \log p(y_i|x_i,\beta) + \lambda[(1-\alpha)||\beta||_2^2/2 + \alpha||\beta||_1]$$

where α and λ are the penalty tuning parameters; α controls the elastic net penalty and λ controls the overall magnitude of the penalty term [68]. When α = 1, the solution amounts to the LASSO regression problem and the coefficients are penalized by the ℓ_1 norm of the coefficients vector whereas when α = 0 it yields the solution to the ridge regression problem involving only ℓ_2 penalization. Otherwise, the elastic net assumes $\alpha \in (0,1)$ giving it flexibility over the former counterparts in that it allows for variable selection potentially leading to a parsimonious model (unlike ridge regression) and will not arbitrarily remove all variables except one in a group of correlated explanatory variables (unlike the LASSO model). In this application, we used known risk factors of FI (dichotomized as: age ≥ 60 years, household receipt of SNAP benefits in the last 12 months, and household size ≥ 5) in addition to the outcome of FI itself to implement the penalized regression models for deriving dietary patterns associated with those outcomes [35,36]. These patterns would subsequently be named accordingly (FI, Age, SNAP, and Household Size) for the remainder of the analysis. The models were fit using the *glmnet* package in R on data from cancer survivors in subsample A. Optimal combinations of α and λ for each of the outcome models were ascertained via 10-fold cross-validation and iterating over a grid of α values ranging from 0 to 1 (inclusive of LASSO and ridge regression). The set of coefficients linked to the combination of α and λ that minimized the deviance was retained as the ultimate set of parameters for a given outcome. Food group explanatory variables were centered and standardized prior to fitting these models in the same manner as the PCA procedure. Given that the *glmnet* software uses a model-based approach rather than a design-based approach for ascertaining model estimates and variances, we weighted the analysis using normalized weights [60].

References

1. Coleman-Jensen, A.; Rabbitt, M.P.; Gregory, C. a; Singh, A. Household Food Security in the United States in 2019 Available online: http://www.ers.usda.gov/publications/pub-details/?pubid=99281 (accessed on 12 August 2021).
2. Charkhchi, P.; Fazeli Dehkordy, S.; Carlos, R.C. Housing and Food Insecurity, Care Access, and Health Status Among the Chronically Ill: An Analysis of the Behavioral Risk Factor Surveillance System. *J. Gen. Intern. Med.* **2018**, *33*, 644–650, doi:10.1007/s11606-017-4255-z.
3. Kudre, D.; Chen, Z.; Richard, A.; Cabaset, S.; Dehler, A.; Schmid, M.; Rohrmann, S. Multidisciplinary Outpatient Cancer Rehabilitation Can Improve Cancer Patients' Physical and Psychosocial Status—a Systematic Review. *Curr. Oncol. Rep.* **2020**, *22*, 122, doi:10.1007/s11912-020-00979-8.

4. Mariotto, A.B.; Enewold, L.; Zhao, J.; Zeruto, C.A.; Yabroff, K.R. Medical Care Costs Associated with Cancer Survivorship in the United States. *Cancer Epidemiol. Biomarkers Prev.* **2020**, *29*, 1304–1312, doi:10.1158/1055-9965.EPI-19-1534.
5. Han, X.; Zhao, J.; Zheng, Z.; de Moor, J.S.; Virgo, K.S.; Yabroff, K.R. Medical Financial Hardship Intensity and Financial Sacrifice Associated with Cancer in the United States. *Cancer Epidemiol. Biomark. Prev. Publ. Am. Assoc. Cancer Res. Cosponsored Am. Soc. Prev. Oncol.* **2020**, *29*, 308–317, doi:10.1158/1055-9965.EPI-19-0460.
6. Simmons, L.A.; Modesitt, S.C.; Brody, A.C.; Leggin, A.B. Food Insecurity Among Cancer Patients in Kentucky: A Pilot Study. *J. Oncol. Pract.* **2006**, *2*, 7.
7. NCI's Dictionary of Cancer Terms.
8. Gany, F.; Leng, J.; Ramirez, J.; Phillips, S.; Aragones, A.; Roberts, N.; Mujawar, M.I.; Costas-Muñiz, R. Health-Related Quality of Life of Food-Insecure Ethnic Minority Patients With Cancer. *J. Oncol. Pract.* **2015**, *11*, 396–402, doi:10.1200/JOP.2015.003962.
9. The American Institute for Cancer Research/World Cancer Research Fund *Diet, Nutrition, Physical Activity and Cancer: A Global Perspective*; 3rd ed.;
10. Thompson, K.L.; Elliott, L.; Fuchs-Tarlovsky, V.; Levin, R.M.; Voss, A.C.; Piemonte, T. Oncology Evidence-Based Nutrition Practice Guideline for Adults. *J. Acad. Nutr. Diet.* **2017**, *117*, 297-310.e47, doi:10.1016/j.jand.2016.05.010.
11. Pekmezi, D.W.; Demark-Wahnefried, W. Updated Evidence in Support of Diet and Exercise Interventions in Cancer Survivors. *Acta Oncol.* **2011**, *50*, 167–178, doi:10.3109/0284186X.2010.529822.
12. Balhareth, A.; Aldossary, M.Y.; McNamara, D. Impact of Physical Activity and Diet on Colorectal Cancer Survivors' Quality of Life: A Systematic Review. *World J. Surg. Oncol.* **2019**, *17*, 153, doi:10.1186/s12957-019-1697-2.
13. Wayne, S.J.; Baumgartner, K.; Baumgartner, R.N.; Bernstein, L.; Bowen, D.J.; Ballard-Barbash, R. Diet Quality Is Directly Associated with Quality of Life in Breast Cancer Survivors. *Breast Cancer Res. Treat.* **2006**, *96*, 227–232, doi:10.1007/s10549-005-9018-6.
14. Zhang, F.; Tapera, T.M.; Gou, J. Application of a New Dietary Pattern Analysis Method in Nutritional Epidemiology. *BMC Med. Res. Methodol.* **2018**, *18*, 119, doi:10.1186/s12874-018-0585-8.
15. McEligot, A.J.; Poynor, V.; Sharma, R.; Panangadan, A. Logistic LASSO Regression for Dietary Intakes and Breast Cancer. *Nutrients* **2020**, *12*, 2652, doi:10.3390/nu12092652.
16. Curtin, L.R.; Mohadjer, L.K.; Dohrmann, S.M.; Kruszon-Moran, D.; Mirel, L.B.; Carroll, M.D.; Hirsch, R.; Burt, V.L.; Johnson, C.L. National Health and Nutrition Examination Survey: Sample Design, 2007-2010. *Vital Health Stat. 2.* **2013**, 1–23.
17. About the National Health and Nutrition Examination Survey. https://www.cdc.gov/nchs/nhanes/about_nhanes.htm
18. Yaghjyan, L.; Wijayabahu, A.T.; Egan, K.M. RE: The Association Between Dietary Quality and Overall and Cancer-Specific Mortality Among Cancer Survivors, NHANES III. *JNCI Cancer Spectr.* **2018**, *2*, pky044, doi:10.1093/jncics/pky044.
19. Wolfe, A.M.; Lee, J.A.; Laurson, K.R. Socioeconomic Status and Physical Fitness in Youth: Findings from the NHANES National Youth Fitness Survey. *J. Sports Sci.* **2020**, *38*, 534–541, doi:10.1080/02640414.2020.1713688.
20. Agarwal, S. The Association of Active and Passive Smoking with Peripheral Arterial Disease: Results from NHANES 1999–2004. *Angiology* **2009**, *60*, 335–345, doi:10.1177/0003319708330526.
21. Dietary Guidelines Advisory Committee; OverDrive, I. *Dietary Guidelines for Americans 2015-2020*; 2016; ISBN 978-0-16-093465-0.

22. Zhao, H.; Pan, Y.; Wang, C.; Guo, Y.; Yao, N.; Wang, H.; Li, B. The Effects of Metal Exposures on Charlson Comorbidity Index Using Zero-Inflated Negative Binomial Regression Model: NHANES 2011–2016. *Biol. Trace Elem. Res.* **2021**, *199*, 2104–2111, doi:10.1007/s12011-020-02331-4.
23. Tucker, L.A. Physical Activity and Telomere Length in U.S. Men and Women: An NHANES Investigation. *Prev. Med.* **2017**, *100*, 145–151, doi:10.1016/j.ypmed.2017.04.027.
24. Blanton, C.A.; Moshfegh, A.J.; Baer, D.J.; Kretsch, M.J. The USDA Automated Multiple-Pass Method Accurately Estimates Group Total Energy and Nutrient Intake. *J. Nutr.* **2006**, *136*, 2594–2599, doi:10.1093/jn/136.10.2594.
25. Moshfegh, A.J.; Rhodes, D.G.; Baer, D.J.; Murayi, T.; Clemens, J.C.; Rumpler, W.V.; Paul, D.R.; Sebastian, R.S.; Kuczynski, K.J.; Ingwersen, L.A.; et al. The US Department of Agriculture Automated Multiple-Pass Method Reduces Bias in the Collection of Energy Intakes. *Am. J. Clin. Nutr.* **2008**, *88*, 324–332, doi:10.1093/ajcn/88.2.324.
26. Jovanovic, C.E.S.; Hoelscher, D.M.; Chen, B.; Ranjit, N.; van den Berg, A.E. The Associations of Plant-Based Food and Metabolic Syndrome Using NHANES 2015–16 Data. *J. Public Health* **2022**, fdab403, doi:10.1093/pubmed/fdab403.
27. Moore, C.; Murphy, M.M.; Keast, D.R.; Holick, M.F. Vitamin D Intake in the United States. *J. Am. Diet. Assoc.* **2004**, *104*, 980–983, doi:10.1016/j.jada.2004.03.028.
28. Montville, J.B.; Ahuja, J.K.C.; Martin, C.L.; Heendeniya, K.Y.; Omolewa-Tomobi, G.; Steinfeldt, L.C.; Anand, J.; Adler, M.E.; LaComb, R.P.; Moshfegh, A. USDA Food and Nutrient Database for Dietary Studies (FNDDS), 5.0. *Procedia Food Sci.* **2013**, *2*, 99–112, doi:10.1016/j.profoo.2013.04.016.
29. Bowman SA, Clemens JC, Friday JE, and Moshfegh AJ. 2020. Food Patterns Equivalents Database 2017-2018: Methodology and User Guide [Online]. Food Surveys Research Group, Beltsville Human Nutrition Research Center, Agricultural Research Service, U.S. Department of Agriculture, Beltsville, Maryland. October 2020. Available at: Http://Www.Ars.Usda.Gov/Nea/Bhnrc/Fsrg.
30. Bowman SA, Friday JE, Moshfegh A. (2008). MyPyramid Equivalents Database, 2.0 for USDA Survey Foods, 2003-2004 [Online] Food Surveys Research Group. Beltsville Human Nutrition Research Center, Agricultural Research Service, U.S. Department of Agriculture, Beltsville, MD. Available at: Http://Www.Ars.Usda.Gov/Ba/Bhnrc/Fsrg.
31. Willett, W.C.; Howe, G.R.; Kushi, L.H. Adjustment for Total Energy Intake in Epidemiologic Studies. *Am. J. Clin. Nutr.* **1997**, *65*, 1220S-1228S, doi:10.1093/ajcn/65.4.1220S.
32. Petrova, D.; Catena, A.; Rodríguez-Barranco, M.; Redondo-Sánchez, D.; Bayo-Lozano, E.; Garcia-Retamero, R.; Jiménez-Moleón, J.-J.; Sánchez, M.-J. Physical Comorbidities and Depression in Recent and Long-Term Adult Cancer Survivors: NHANES 2007–2018. *Cancers* **2021**, *13*, 3368, doi:10.3390/cancers13133368.
33. Bickel, G.; Nord, M.; Price, C.; Hamilton, W.; Cook, J. Guide to Measuring Household Food Security 2000.
34. U.S. Household Food Insecurity Survey Module: Three-Stage Design, With Screeners. https://www.ers.usda.gov/media/8271/hh2012.pdf
35. Lee, J.S.; Frongillo, E.A. Nutritional and Health Consequences Are Associated with Food Insecurity among U.S. Elderly Persons. *J. Nutr.* **2001**, *131*, 1503–1509, doi:10.1093/jn/131.5.1503.
36. Kohn, M.J.; Bell, J.F.; Grow, H.M.G.; Chan, G. Food Insecurity, Food Assistance and Weight Status in US Youth: New Evidence from NHANES 2007-08: Food Insecurity, Assistance and Weight. *Pediatr. Obes.* **2014**, *9*, 155–166, doi:10.1111/j.2047-6310.2012.00143.x.
37. Clifford Johnson; Paulose-Ram, R.; Ogden, C.L.; Carroll, M.; Kruszan-Moran, D.; Dohrmann, S.; Curtin, L. National Health and Nutrition Examination Survey. Analytics Guidelines, 1999-2010. *Vital Health Stat. Ser. 2* **2013**, 1–16.

38. Goldberg, R. PROC FACTOR: How to Interpret the Output of a Realworld Example.
39. Tsuruga, K.; Sugawara, N.; Sato, Y.; Saito, M.; Furukori, H.; Nakagami, T.; Nakamura, K.; Takahashi, I.; Nakaji, S.; Yasui-Furukori, N. Dietary Patterns and Schizophrenia: A Comparison with Healthy Controls. *Neuropsychiatr. Dis. Treat.* **2015**, *11*, 1115–1120, doi:10.2147/NDT.S74760.
40. Kopp, W. How Western Diet And Lifestyle Drive The Pandemic Of Obesity And Civilization Diseases. *Diabetes Metab. Syndr. Obes. Targets Ther.* **2019**, *12*, 2221–2236, doi:10.2147/DMSO.S216791.
41. Azzam, A. Is the World Converging to a 'Western Diet'? *Public Health Nutr.* **2021**, *24*, 309–317, doi:10.1017/S136898002000350X.
42. Arthur, A.E.; Peterson, K.E.; Rozek, L.S.; Taylor, J.M.G.; Light, E.; Chepeha, D.B.; Hébert, J.R.; Terrell, J.E.; Wolf, G.T.; Duffy, S.A.; et al. Pretreatment Dietary Patterns, Weight Status, and Head and Neck Squamous Cell Carcinoma Prognosis. *Am. J. Clin. Nutr.* **2013**, *97*, 360–368, doi:10.3945/ajcn.112.044859.
43. El Zein, A.; Colby, S.E.; Zhou, W.; Shelnutt, K.P.; Greene, G.W.; Horacek, T.M.; Olfert, M.D.; Mathews, A.E. Food Insecurity Is Associated with Increased Risk of Obesity in US College Students. *Curr. Dev. Nutr.* **2020**, *4*, nzaa120, doi:10.1093/cdn/nzaa120.
44. Larson, N.; Laska, M.N.; Neumark-Sztainer, D. Food Insecurity, Diet Quality, Home Food Availability, and Health Risk Behaviors Among Emerging Adults: Findings From the EAT 2010–2018 Study. *Am. J. Public Health* **2020**, *110*, 1422–1428, doi:10.2105/AJPH.2020.305783.
45. Eicher-Miller, H.A.; Zhao, Y. Evidence for the Age-Specific Relationship of Food Insecurity and Key Dietary Outcomes among US Children and Adolescents. *Nutr. Res. Rev.* **2018**, *31*, 98–113, doi:10.1017/S0954422417000245.
46. Shi, Y.; Davies, A.; Allman-Farinelli, M. The Association Between Food Insecurity and Dietary Outcomes in University Students: A Systematic Review. *J. Acad. Nutr. Diet.* **2021**, *121*, 2475-2500.e1, doi:10.1016/j.jand.2021.07.015.
47. Faught, E.L.; Williams, P.L.; Willows, N.D.; Asbridge, M.; Veugelers, P.J. The Association between Food Insecurity and Academic Achievement in Canadian School-Aged Children. *Public Health Nutr.* **2017**, *20*, 2778–2785, doi:10.1017/S1368980017001562.
48. Mello, J.A.; Gans, K.M.; Risica, P.M.; Kirtania, U.; Strolla, L.O.; Fournier, L. How Is Food Insecurity Associated with Dietary Behaviors? An Analysis with Low-Income, Ethnically Diverse Participants in a Nutrition Intervention Study. *J. Am. Diet. Assoc.* **2010**, *110*, 1906–1911, doi:10.1016/j.jada.2010.09.011.
49. Zhao, J.; Li, Z.; Gao, Q.; Zhao, H.; Chen, S.; Huang, L.; Wang, W.; Wang, T. A Review of Statistical Methods for Dietary Pattern Analysis. *Nutr. J.* **2021**, *20*, 37, doi:10.1186/s12937-021-00692-7.
50. Jung, N.M.; de Bairros, F.S.; Pattussi, M.P.; Pauli, S.; Neutzling, M.B. Gender Differences in the Prevalence of Household Food Insecurity: A Systematic Review and Meta-Analysis. *Public Health Nutr.* **2017**, *20*, 902–916, doi:10.1017/S1368980016002925.
51. Martin, M.A.; Lippert, A.M. Feeding Her Children, but Risking Her Health: The Intersection of Gender, Household Food Insecurity and Obesity. *Soc. Sci. Med.* **2012**, *74*, 1754–1764, doi:10.1016/j.socscimed.2011.11.013.
52. Franklin, B.; Jones, A.; Love, D.; Puckett, S.; Macklin, J.; White-Means, S. Exploring Mediators of Food Insecurity and Obesity: A Review of Recent Literature. *J. Community Health* **2012**, *37*, 253–264, doi:10.1007/s10900-011-9420-4.
53. Lohman, B.J.; Neppl, T.K.; Lee, Y.; Diggs, O.N.; Russell, D. The Association between Household Food Insecurity and Body Mass Index: A Prospective Growth Curve Analysis. *J. Pediatr.* **2018**, *202*, 115-120.e1, doi:10.1016/j.jpeds.2018.05.052.

54. Nettle, D.; Andrews, C.; Bateson, M. Food Insecurity as a Driver of Obesity in Humans: The Insurance Hypothesis. *Behav. Brain Sci.* **2017**, *40*, e105, doi:10.1017/S0140525X16000947.
55. Rasmusson, G.; Lydecker, J.A.; Coffino, J.A.; White, M.A.; Grilo, C.M. Household Food Insecurity Is Associated with Binge-Eating Disorder and Obesity. *Int. J. Eat. Disord.* **2019**, *52*, 28–35, doi:10.1002/eat.22990.
56. Trego, M.L.; Baba, Z.M.; DiSantis, K.I.; Longacre, M.L. Food Insecurity among Adult Cancer Survivors in the United States. *J. Cancer Surviv.* **2019**, *13*, 641–652, doi:10.1007/s11764-019-00783-9.
57. Patel, K.G.; Borno, H.T.; Seligman, H.K. Food Insecurity Screening: A Missing Piece in Cancer Management. *Cancer* **2019**, *125*, 3494–3501, doi:10.1002/cncr.32291.
58. Gany, F.; Lee, T.; Ramirez, J.; Massie, D.; Moran, A.; Crist, M.; McNish, T.; Winkel, G.; Leng, J.C.F. Do Our Patients Have Enough to Eat?: Food Insecurity among Urban Low-Income Cancer Patients. *J. Health Care Poor Underserved* **2014**, *25*, 1153–1168, doi:10.1353/hpu.2014.0145.
59. Gany, F.; Bari, S.; Crist, M.; Moran, A.; Rastogi, N.; Leng, J. Food Insecurity: Limitations of Emergency Food Resources for Our Patients. *J. Urban Health Bull. N. Y. Acad. Med.* **2013**, *90*, 552–558, doi:10.1007/s11524-012-9750-2.
60. McConville, K. Improved Estimation for Complex Surveys Using Modern Regression Techniques. *Colo. State Univ. Fort Collins CO USA*.
61. McConville, K.S.; Breidt, F.J.; Lee, T.C.M.; Moisen, G.G. Model-Assisted Survey Regression Estimation with the Lasso. *J. Surv. Stat. Methodol.* **2017**, *5*, 131–158, doi:10.1093/jssam/smw041.
62. Fransen, H.P.; May, A.M.; Stricker, M.D.; Boer, J.M.A.; Hennig, C.; Rosseel, Y.; Ocké, M.C.; Peeters, P.H.M.; Beulens, J.W.J. A Posteriori Dietary Patterns: How Many Patterns to Retain? *J. Nutr.* **2014**, *144*, 1274–1282, doi:10.3945/jn.113.188680.
63. Kant, A.K. Dietary Patterns and Health Outcomes. *J. Am. Diet. Assoc.* **2004**, *104*, 615–635, doi:10.1016/j.jada.2004.01.010.
64. Abdi, H.; Williams, L.J. Principal Component Analysis: Principal Component Analysis. *Wiley Interdiscip. Rev. Comput. Stat.* **2010**, *2*, 433–459, doi:10.1002/wics.101.
65. Lumley, T. Analysis of Complex Survey Samples. *J. Stat. Softw.* **2004**, *9*, doi:10.18637/jss.v009.i08.
66. Friedman, J.; Hastie, T.; Tibshirani, R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J. Stat. Softw.* **2010**, *33*, 1–22.
67. Tibshirani, R. Regression Shrinkage and Selection Via the Lasso. *J. R. Stat. Soc. Ser. B Methodol.* **1996**, *58*, 267–288, doi:10.1111/j.2517-6161.1996.tb02080.x.
68. Hastie, T.; Qian, J.; Tay, K. An Introduction to `glmnet` 2021.