**Methods**

*Study Population*

The goal of this analysis was to characterize the associations between prevalent dietary patterns in the food insecure cancer survivor population and the risk of mortality among the broader cancer survivor population and the food insecure cancer survivor population. We employed data from ten consecutive biennial cycles (1999-2018) from the National Health and Nutrition Examination Study (NHANES). The NHANES is a biennial cross-sectional study implemented by the Centers for Disease Control and Prevention (CDC) and the National Center for Health Statistics and samples civilian and non-institutionalized community dwellers in the United States through 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 implements a series of surveys that span numerous facets of health and lifestyle. In addition, a subsample is selected to participate in a series of 24-hour recalls to gauge dietary intake. Finally, subjects may also be selected for a subsample that undergoes a medical examination in the Mobile Examination Center consisting of a number of physical and anthropomorphic measurements. All subjects provided informed and written consent and all study protocols were approved by the NCHS Ethics Review Board [1]. 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 [2,3]. The survey contains ten items for households with only adults and an additional eight items completed by subjects living in households with children. 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 not having a significant history of cancer given that prognosis and benign course of these malignancies that may otherwise bias the sample [4].

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 [5]. 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) [6,7]. Nutrient intake data were estimated by referencing the Food and Nutrient Database for Dietary Studies [8]. Dietary intake and nutrient data were averaged across both 24-hour recalls as previously described [5,9,10]. 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 [5]. A multivariate density model approach was used for adjusting food and nutrient intake levels for total energy consumption [11]. 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 [12,13]. See Maino Vieytes et al. (2022) for a detailed procedural description and a discussion about the component retention process for the PCA [5].

*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 [14,15]. 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 [16–18]. 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 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 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 [19–21]. Health-related covariates included a Charlson Comorbidity Index score (adapted for NHANES) and body mass index measured during the physical examination (BMI—kilograms/m2) [5,22]. 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 [23,24]. 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*

Descriptive statistics were generated for the explanatory, response, and covariate variables described above. We also examined the correlations between the extracted diet quality measures and the 26 food groups used in the extraction process to evaluate the extent of how food groups contributed to each of those measures. 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.

In equation 1 we specify the diet quality index, , using dummy variables, , that indicate the subject’s membership in one of the quintiles. In equation 2 we conduct a trend test by assigning 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. Additionally, all models included terms () for covariates ().