1.a. **Full Title:** Dietary patterns and cancer incidence in the ARIC cohort

b. **Abbreviated Title (Length 26 characters):** Dietary patterns and cancer

2. **Writing Group:**
   Patrick Bradshaw
   June Stevens
   Other interested investigators

I, the first author, confirm that all the coauthors have given their approval for this manuscript proposal. _PTB_ [please confirm with your initials electronically or in writing]

First author: Patrick T. Bradshaw, Ph.D.
Address: University of North Carolina
         Department of Epidemiology
         CB #7435
         Chapel Hill, NC 27599-7435

Phone: 919/360-8056       Fax: 919/966-7215
E-mail: patrickb@email.unc.edu

**ARIC author** to be contacted if there are questions about the manuscript and the first author does not respond or cannot be located (this must be an ARIC investigator).

Name: June Stevens, Ph.D.
Address: University of North Carolina
         Department of Nutrition, CB #7461
         Chapel Hill, North Carolina 27599-7461

Phone: 919/966-7218       Fax: 919/966-7215
E-mail: june_stevens@unc.edu

3. **Timeline:**

   We are currently seeking additional funding for this analysis. If funding is not obtained in the near future analysis will still be completed before this manuscript proposal expires.
4. Rationale:

Dietary behaviors have been linked to risk for a number of cancers, especially those with a strong connection to obesity, including those of the breast, colon and esophagus [1, 2]. To date, most research on diet and cancer has focused on analyses of individual foods, food groups or nutrients. This approach is problematic due to the high correlation between reported food items and issues with multiple comparisons in the analysis of many factors simultaneously. Recently, the analysis of dietary patterns has been advocated as an alternative to the standard paradigm [3]. The objective of this approach is to reduce a series of highly correlated dietary variables (food intakes) into a set of distinct patterns that represent homogeneous behaviors that may be related to disease status.

Common methods of dietary pattern analysis are classified into \textit{a priori} or \textit{a posteriori} techniques. \textit{A priori} dietary patterns, such as indices of healthy eating, measure each individual’s compliance with a pre-defined healthy diet. In contrast, \textit{a posteriori} dietary patterns, such as factor analysis or cluster analysis, organize subjects according to observed patterns of dietary intakes, which do not invoke any set definition of healthy eating. One popular \textit{a posteriori} approach is factor analysis, a procedure that utilizes the correlation between food items to derive a set of latent dietary behaviors, with each subject exhibiting some degree of each of them. A limitation of such methods is their ignorance of the biological relationship between diet and disease. However, recently this approach has been extended to allow incorporation of biologic information by the method of reduced rank regression (RRR) [4]. Reduced rank regression derives dietary factors by utilizing a set of intermediates, referred to as response variables, thought to be related to both diet (as a consequence) and disease (as a cause). This approach is viewed as a combination of the \textit{a priori} and \textit{a posteriori} approaches since it relies on prior biologic knowledge to derive empirically observed dietary patterns.

\textbf{Dietary Patterns and Cancer Risk}

A number of studies have examined the relationship between dietary patterns and incident cancers. Most of these have been case-control studies with the vast majority focusing on \textit{a posteriori} dietary patterns through the use of factor analysis or principal component analysis of intakes assessed through food frequency questionnaire. A recent review of the literature on dietary patterns and breast cancer identified 19 studies on this subject published between 1995 and 2008 [5]. Thirteen of these found at least one dietary pattern associated with breast cancer risk with various definitions of “Western” dietary patterns being associated with an increase in risk, while several “prudent” patterns were inversely associated with risk. A similar review of dietary patterns and colorectal cancer identified 32 studies with patterns of a “Western” diet (typically high in red and processed meats and refined carbohydrates) being associated with an increase in colorectal cancer or adenomas while more prudent diets (high in fruit, vegetables and lean meats) were associated with fewer cases [6]. The associations between dietary patterns and cancers other than breast and colorectal have been much less studied, however reports have shown associations for lung cancer [7], head and neck cancer [8-12], esophageal cancer [13, 14] and prostate cancer [15].

Although the previous studies yielded intriguing findings, most were limited by the use of purely empirically derived dietary patterns, which, as described above, ignore any knowledge of the biological relationship between diet and disease. Only two studies in the review of the breast...
cancer literature utilized RRR, one deriving a “high glycemic load” pattern, and one high in consumption of animal fats and processed meats, which were both positively associated with breast cancer risk [5]. The study of dietary patterns and cancer would benefit greatly by leveraging our understanding of the physiological effects of diet.

One potential link between diet and cancer is likely through its impact on insulin resistance [16] which is believed to be the underlying cause of the metabolic syndrome, a cluster of physiologic abnormalities thought to result from obesity-induced hyperinsulinemia and inflammation [17, 18]. Insulin resistance is considered to be a contributing factor in carcinogenesis [19] and the metabolic syndrome has been shown to be associated with many chronic diseases, including cancer [20]. Additionally, cardiometabolic abnormalities including elevated blood pressure and excess adipose tissue may influence cancer risk independently of insulin’s mitogenic effects [21-24]. Since diet has been shown to be a factor in the development of the metabolic syndrome [25] utilizing its components as the basis for deriving dietary patterns through reduced rank regression could prove useful.

**Related Studies in ARIC**
Specific to the ARIC cohort, metabolic syndrome (the presence of at least 3 of: low HDL cholesterol, elevated triglycerides, elevated waist circumference, high blood pressure and elevated fasting blood glucose) has been linked to an increased risk of colorectal cancer among men [26], while inversely related to prostate cancer [27]. Low high density lipoprotein (HDL) cholesterol levels, one of the components of metabolic syndrome, has also been linked to increased risk of lung cancer [29] as well as increased risk of breast cancer among premenopausal women [28]. Furthermore, a recent analysis using ARIC data has shown dietary behaviors, including a “Western” dietary pattern (frequent consumption of refined grains, processed meat, fried foods, and red meat), to be associated with risk of developing metabolic syndrome [25]. Given these previous findings, the use of dietary patterns derived through information based on the components of the metabolic syndrome could provide important insight into the complex relationship between diet and cancer.

**Study Innovation**
This study will be the first to our knowledge to examine the association between empirically-derived dietary patterns based on metabolic characteristics and cancer risk and therefore represents an important contribution to the literature. The prospective nature of the ARIC cohort, combined with its repeated physiological and anthropometric assessments makes it an ideal data source for this analysis. Furthermore, this will be the first study in the ARIC cohort to examine the association between diet and cancer incidence.

5. **Main Hypothesis/Study Questions:**

   **Aim 1:** What are the dietary patterns associated with cardiometabolic characteristics (components of the metabolic syndrome) at each ARIC visit for subjects in the cancer cohort?
Aim 2: What are the associations between the latent dietary patterns derived in aim 1 and risk of individual cancers (including breast, colon, endometrial, prostate, esophageal, pancreatic and lung) over the follow-up?

6. Design and analysis (study design, inclusion/exclusion, outcome and other variables of interest with specific reference to the time of their collection, summary of data analysis, and any anticipated methodologic limitations or challenges if present).

Study Design: Aim 1 will employ a cross-sectional study design at visits 1 and visit 3 (the two visits with dietary assessments). Aim 2 will be a prospective cohort study and will employ all data from baseline and the three follow-up visits among those individuals who meet eligibility requirements.

Inclusion/Exclusion Criteria: For both aims we will exclude those subjects with implausible total energy intake, and missing dietary, physical activity, or other covariate data. If, upon undertaking the analysis, the proportion of missing data for these or other variables is concerning then formal missing data methods, such as multiple imputation, will be employed.

Outcome variables: Incident cancers available through the ARIC cohort, including cancer of the breast, colon, endometrial, prostate, esophageal, pancreatic, lung. Each cancer will be analyzed separately.

Covariates: The primary exposures for the analysis in Aim 2 will be dietary patterns derived through RRR of food intakes on components of the metabolic syndrome (systolic and diastolic blood pressure, fasting blood glucose, serum high density lipoprotein (HDL) level, serum triglyceride (TG) level, waist circumference in centimeters). The metabolic syndrome components will be treated continuously for the RRR analysis, as is required for this procedure. The latent dietary patterns derived from this method will be categorized according to percentiles (quartiles or quintiles) for the regression analysis. Potential confounders for this analysis include: age, sex, race-ethnicity, education level, menopausal status, hormone use, smoking status, alcohol use, leisure time physical activity, total energy intake, and family history of diabetes or cardiovascular disease.

Statistical Analysis: For Aim 1 we will obtain dietary patterns through RRR of food-item intakes on indicators of metabolic health, as described above. For Aim 2, we will use these dietary patterns, categorized as described above, to estimate unadjusted (no confounder adjustment) and multivariate adjusted Cox proportional hazards models (dietary pattern adjusted for relevant covariates) with the outcome time to incident cancer (or censoring) for each cancer type (breast, colon, endometrial, prostate, esophageal). Since dietary assessments were conducted at visits 1 and 3, we will treat the dietary patterns as time-varying covariates, with the values updated at these visits. Test for trend will be calculated utilizing the uncategorized variable. For cancers with sufficient numbers of events to explore effect modification, we will compare the interaction between dietary patterns and potential effect modifiers (such as body mass index (BMI) or physical activity level) using the likelihood ratio test for multiplicative interaction with a significance level of 0.05.
7.a. Will the data be used for non-CVD analysis in this manuscript?  _X_ Yes  ____ No

   b. If Yes, is the author aware that the file ICTDER03 must be used to exclude persons with a value RES_OTH = “CVD Research” for non-DNA analysis, and for DNA analysis RES_DNA = “CVD Research” would be used?  _X_ Yes  ____ No
   (This file ICTDER03 has been distributed to ARIC PIs, and contains the responses to consent updates related to stored sample use for research.)

8.a. Will the DNA data be used in this manuscript?  ____ Yes  ___X_ No

8.b. If yes, is the author aware that either DNA data distributed by the Coordinating Center must be used, or the file ICTDER03 must be used to exclude those with value RES_DNA = “No use/storage DNA”?  ____ Yes  ____ No

9. The lead author of this manuscript proposal has reviewed the list of existing ARIC Study manuscript proposals and has found no overlap between this proposal and previously approved manuscript proposals either published or still in active status. ARIC Investigators have access to the publications lists under the Study Members Area of the web site at: http://www.cscc.unc.edu/ARIC/search.php  
   _X_ Yes  _______ No

10. What are the most related manuscript proposals in ARIC (authors are encouraged to contact lead authors of these proposals for comments on the new proposal or collaboration)?

MS #1117: Different methods for deriving dietary patterns and their relation with risk of developing incident coronary heart disease: The ARIC Study.

MS#1117 focuses on comparisons between different dietary pattern methods (a variety of both a priori and a posteriori approaches) and risk of coronary heart disease. We will be focusing our analysis on incident cancers.

MS #1173: Dietary intake and the development of the metabolic syndrome: The ARIC Study.

MS#1173 focused on the association between dietary patterns (derived by factor analysis), food groups and incident metabolic syndrome (as a single entity). Our analysis will utilize the five components of the metabolic syndrome separately as the intermediates (response variables) in the derivation of dietary patterns through reduced rank regression. These patterns will then be related to incident cancer, instead of incident metabolic syndrome as with MS#1173.

11. a. Is this manuscript proposal associated with any ARIC ancillary studies or use any ancillary study data? _X_ Yes  ____ No

11.b. If yes, is the proposal  
   ___  A. primarily the result of an ancillary study (list number* _1995.04_*)
B. primarily based on ARIC data with ancillary data playing a minor role
(usually control variables; list number(s)* __________ __________ __________)

*ancillary studies are listed by number at http://www.cscct.unc.edu/aric/forms/

12. Manuscript preparation is expected to be completed in one to three years. If a
manuscript is not submitted for ARIC review at the end of the 3-years from the date of
the approval, the manuscript proposal will expire.

REFERENCES