1.a. **Full Title**: Metabolomic markers of carbohydrate variant diets as they relate to mortality

b. **Abbreviated Title (Length 26 characters)**: Metabolomics and carbohydrates

2. **Writing Group**:
   Writing group members: Sara B. Seidelmann, Bing Yu, Brian Claggett, Casey M. Rebholz, Lyn M. Steffen, Susan Cheng, Eric Boerwinkle, Scott D. Solomon, others welcome

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

**First author**: Sara B. Seidelmann  
Address: Brigham and Women’s Hospital  
Cardiovascular Division  
75 Francis Street, ASB 1-L1-037B  
Boston, MA 02115

Phone: 917-593-7506  
Fax: 617-582-6056  
E-mail: sseidelmann@bwh.harvard.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: Scott D. Solomon  
Address: Brigham and Women’s Hospital  
Cardiovascular Division  
75 Francis Street  
Boston, MA 02115

Phone: 857-307-1960  
Fax: 857-307-1944  
E-mail: ssolomon@rics.bwh.harvard.edu

3. **Timeline**: Analysis will begin following proposal approval with the aim of completing the analysis and associated manuscript(s) within 1 year of data availability.

4. **Rationale**: 
Low carbohydrate (LC) diets that exchange carbohydrates for a greater intake of protein and/or fat have gained in popularity due to their ability to induce short-term weight loss\textsuperscript{1-4} despite limited and conflicting data regarding their longer-term effects on health outcomes.\textsuperscript{5-9} To better understand the relationship between dietary carbohydrate intake and mortality, our group previously utilized data from the ARIC study (currently under review for publication; ARIC manuscript proposal #2844) which showed that both low carbohydrate (<40 percent of energy from carbohydrate) and high carbohydrate (>70 percent of energy from carbohydrate) consumption were associated with increased mortality risk and shorter residual lifespan, reflecting a U-shaped relationship between carbohydrate intake and mortality. Our primary results were confirmed in the context of data from other North American, European, Asian and multinational cohorts, combined as part of a meta-analysis. When evaluating total carbohydrate without regard to specific food source, both high (>70%) and low (<40%) percent energy from carbohydrates were associated with increased mortality, with minimal risk detected between 50-55%. Importantly, mortality risks conferred by low carbohydrate diets varied depended on whether carbohydrates were replaced by animal versus plant-based foods: mortality increased when carbohydrates were exchanged for animal-derived fat or protein and mortality decreased when the substitutions were plant-based. To follow up on our prior work, it is of interest to explore the possible mechanisms by which these various LC dietary patterns may relate to longevity outcomes by examining their association with metabolomic markers in the ARIC population.

5. Main Hypothesis/Study Questions:
We aim to understand which metabolites are most strongly associated with carbohydrate variant diets (total- animal- and plant- LC diets) and how they relate to mortality.

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).

Eligibility: Study participants with serum metabolite data from visit 1 (approximately 4,000 African-American and European-American participants)

Exclusion: Participants without complete dietary information or with extreme caloric intake (defined as <600 kcal or >4200 kcal per day for men and <500 kcal or >3600 kcal per day for women) will be excluded.

Exposures: Nutrition data from Visit 1 will be utilized: percentage of total kcal from carbohydrate, fat (animal, vegetable), and protein (animal and vegetable).

Specifically, we will create 3 scores:
1) Total Low Carbohydrate score (inverse of % energy for carbohydrate)
2) Low Carbohydrate Animal score--Animal- and plant-based scores will be created by dividing subjects into deciles for either animal- or plant derived fat and protein, and carbohydrate intake, expressed as a percentage of energy as previously described.\textsuperscript{10,11} For carbohydrate, subjects in
the lowest decile will receive 10 points, while subjects in the highest decile will receive 1 point. The order will be reversed for animal or plant-derived fat and protein, so that the highest score represents LC and high animal- or plant-derived fat and protein intake.

3) Low Carbohydrate Plant score

Outcomes:
Primary Outcomes--The outcomes studied will be serum metabolites and death. Nontargeted metabolomic analyses of fasting serum samples from ARIC participants at Visit 1 Metabolon, Inc. (Durham, NC, USA) will be used. All-cause death occurring from Visit 1 until 2016 (or most current) will be considered.
Secondary Outcomes—Type of Mortality: Cardiovascular Death and Events: Heart Failure, Stroke, Myocardial Infarction, and Non-CV death

Analysis Plan—Data processing: We will assess how much “missing” data is present and whether this missingness represents true “missing” versus “undetectably low”. If missing represents low, then we will use half the lower limit of detection. We will check to see if the metabolite data needs to be transformed (assess normality). Likely, a log-transformation will be necessary. Next, we will rescale the log-transformed data so that all beta’s are reported “per SD” of the analyte. We will use age and sex as covariates in all analyses. We will perform linear regression analysis for each of the 245 analytes with 1) low carbohydrate score, 2) LC Animal Score 3) LC Plant Score. We will create a list of analytes that are statistically significant after Bonferroni correction (p=0.05/245) for each of the 3 scores with beta coefficient and p-values. In order to understand the relative importance of the analytes identified in predicting the score, for each of the 3 scores separately, we will place those analytes into a single multivariate regression model with the score as the outcome variable. We will perform a forward stepwise selection model with p-value threshold of 0.05. We will report coefficients, Z-scores, and p-values. For example, if 6 analytes are correlated with the animal score then we will perform: First, regress animal score A1 A2 A3 A4 A5 A6 Next, we will perform a forward stepwise selection model with p-value threshold of 0.05. Then we will report coefficients, Z-scores and p-values for A1, A2, A3, A4, A5 and A6 as well as the global p-value. We will repeat above for each of 4 macronutrient components that comprise the “low-carb” scores so that we may be able to further understand their contribution to each score (% animal protein, % vegetable protein, percent animal fat, percent vegetable fat). All analyses will be stratified by race and then combined in meta-analysis.

For the top metabolite candidates identified above that are related to the 3 low carbohydrate scores, we will then assess their relationship with mortality. Cox regression analysis will be performed relating the analytes with all-cause death. Covariates will include age, sex, race, BMI, history of diabetes, physical activity, education, and tobacco use.
Visit 1 will serve as baseline for these analyses (the time at which both the food frequency questionnaire and metabolomics data were obtained).

Anticipated methodologic limitations or challenges if present: Since there is a significant risk for false-positive results based on the large number of tests proposed, we will perform Bonferroni correction as detailed above.

7.a. Will the data be used for non-CVD analysis in this manuscript? _____ Yes  __x__ 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? _____ Yes    _____ No
(This file ICTDER 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    ____ 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)?
The proposals below studied metabolomics as it relates to several disease states and dietary intake. None of them have specifically looked at metabolomics in relation to carbohydrate intake or mortality. Many were limited to African Americans.


#1918: Associations of the human metabolome with blood pressure, prevalent, and incident hypertension among African Americans in the Atherosclerosis Risk in Communities study (lead author: Yan Zheng)

#1847: Role of the human metabolome in incident heart failure etiology among African Americans in the Atherosclerosis Risk in Communities Study (lead author: Yan Zheng)

#2756: Serum Metabolomic Profile of Diabetes and Glycemic Biomarkers (lead author: Casey M. Rebholz)
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
   ___x_  A. primarily the result of an ancillary study (list number* __2017.02___)
   ___    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.csc.unc.edu/aric/forms/

12a. 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.

12b. The NIH instituted a Public Access Policy in April, 2008 which ensures that the public has access to the published results of NIH funded research. It is your responsibility to upload manuscripts to PubMed Central whenever the journal does not and be in compliance with this policy. Four files about the public access policy from http://publicaccess.nih.gov/ are posted in http://www.csc.unc.edu/aric/index.php, under Publications, Policies & Forms.
   http://publicaccess.nih.gov/submit_process_journals.htm shows you which journals automatically upload articles to PubMed central.

13. Per Data Use Agreement Addendum, approved manuscripts using CMS data shall be submitted by the Coordinating Center to CMS for informational purposes prior to publication. Approved manuscripts should be sent to Pingping Wu at CC, at pingping_wu@unc.edu. I will be using CMS data in my manuscript ____ Yes _x____ No.

References