

ARIC Manuscript Proposal #3431

PC Reviewed: 7/9/19

Status: _____

Priority: 2

SC Reviewed: _____

Status: _____

Priority: _____

1.a. Full Title: Proteomics, Cigarette Smoking, and Mortality due to Cigarette-smoking-related Cancers

b. Abbreviated Title (Length 26 characters):

2. Writing Group:

Writing group members: Adrienne Tin, Corinne E. Joshi, Anna Prizment, David Couper, Ken Butler, Christie Ballantyne, Ron Hoogeveen, Elizabeth A. Platz, Josef Coresh, others welcome.

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

First author: Adrienne Tin, PhD

Address: Department of Epidemiology
Johns Hopkins Bloomberg School of Public Health
615 N. Wolfe Street, W6017
Baltimore, MD 21205
Phone: 201-281-9577
atin1@jhu.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: Elizabeth A. Platz, ScD, MPH

Address: Department of Epidemiology
Johns Hopkins Bloomberg School of Public Health
615 N. Wolfe Street, Room E6132
Baltimore, Maryland 21205

Phone: 410-614-9674

Fax: (410) 614-2632

E-mail: eplatz1@jhu.edu

3. Timeline: Analysis using visit 5 SomaLogic protein measures, designed to provide preliminary data for an R01 application to document our use of these data for studies on cancer, will start immediately. We will amend this manuscript proposal to use the visit 2 and 3 SomaLogic protein measures once they are available. The manuscript using visit 2 and 3 protein measures is expected to be submitted in 1 year after the data are available.

4. Rationale:

Cigarette smoking is a well-established causal risk factor of cancers and cancer-related mortality.^{1,2} Levels of some circulating proteins have been found to predict future diagnosis of smoking-related cancers.^{3,4} For example, anti-inflammation markers, such as interleukin 1 receptor antagonist (IL-1RA), and chemokines, such as CXCL5, have been associated with diagnosis of lung cancer.³ These circulating proteins that are associated with diagnosis of smoking-related cancers may also inform smoking-related cancer mortality.

Among ~5,000 participants of the Atherosclerosis Risk in Communities (ARIC) study at visit 5, 4,941 human plasma proteins have been measured using SOMAscan v.4 (SomaLogic, Boulder, CO), an aptamer-based technology.⁵⁻⁷ We propose to conduct a prospective cohort study of circulating protein levels associated with smoking in relation to smoking-related cancer mortality.

While follow-up time is short (2012-2015 through 2017), this analysis using visit 5 proteomics data is designed to provide preliminary data for an R01 application to document our use of these data for studies on cancer. Later when SOMAscan data for visits 2 and 3 are available, we plan to amend this manuscript proposal to address an expanded version of the same question - whether proteins associated with smoking are associated with smoking-associated cancer incidence and mortality with decades of follow-up. Overarching goals of this inquiry are to 1) understand better the spectrum of pathways beyond DNA mutation that may mediate the association between smoking and smoking-associated cancers, and 2) inform why some smokers develop smoking-associated cancers, while others do not.

5. Main Hypothesis/Study Questions:

Some, but not all proteins associated with smoking will be associated with death from smoking-associated cancers.

Smoking status will be analyzed as current/former/never smoker, packyears smoked, and time since quitting among former smokers concurrent with the visit at which the proteins were measured and used in the analysis as baseline.

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).

Described here is for the analysis using the proteins from visit 5.

We propose a prospective cohort study with 4-6 years of follow up from visit 5 through 2017 (or 2018, depending on the availability of the next death file).

Inclusion criteria: Participants with SOMAscan version 4 measures, and covariate data.

Exclusion criteria:

- i) Participants who only consented for CVD research.
- ii) Participants with prevalent cancer at visit 5

- iii) Sensitivity analysis: Participants with prevalent liver and kidney disease, which may affect circulating protein concentrations.
- iv) Sensitivity analysis: Depending on the data on medications and treatment, participants who were on therapies with potentially large effects on circulating protein levels may need to be excluded.

Outcomes:

- i) Protein measures from SOMAscan version 4 at visit 5. This analysis is for identifying proteins that are potential mediators of smoking for smoking-related cancers. We will conduct a cross-sectional analysis using protein measures as outcome and smoking status, packyears, and time since quitting as predictor. We will investigate an appropriate exclusion threshold for proteins that were FLAGGED in at least one plate, and the appropriate transformation of the relative fluorescence units (RFU) of protein levels, such as log2 and/or inverse normal transformation.
- ii) Smoking-related cancer mortality ascertained as the underlying cause on the death certificate (lung, oral cavity and throat, esophagus, larynx, stomach, pancreas, liver, colon and rectum, bladder, kidney, and cervix, and acute myeloid leukemia).

Predictors:

To detect smoking-associated proteins:

- i) Smoking status (current/former/never) as a categorical variable
- ii) Packyears in current and former smokers
- iii) Time since quitting in former smokers

To evaluate associations with smoking-related cancer mortality:

- i) Smoking-associated proteins

Covariates:

Age, sex, race-center, alcohol use, BMI, diabetes, eGFRcr-cys, hypertension and cholesterol medication use. We will also investigate the followings as potential covariates: medications that are known to affect circulating protein levels, cancer family history, vitamin C and E, and beta carotene exposure.

Statistical analysis:

A) Detection of smoking-associated proteins

We will use linear regression to assess the association of smoking status, packyears, and time since quitting with the protein levels as the dependent variable controlling for the covariates.

Model 1: age, sex, race-center

Model 2: alcohol use, BMI, diabetes, hypertension and cholesterol medication use, eGFR

We will set the threshold for selection of proteins to evaluate in relation to smoking-associated cancer mortality of $p < 0.05/\text{number of proteins that are analyzable } (\sim 4,941)$ in the likelihood ratio of test of the smoking variables in Model 2.

B) Assess the association between smoking-associated proteins and smoking-related cancer mortality

We will use Cox regression to assess the association between the selected smoking-associated proteins from A) and smoking-related cancer mortality. We will express the proteins as a continuous variable (with transformation, if appropriate). Covariates will include age, sex, race-center, alcohol use, BMI, diabetes, hypertension and cholesterol medication use, hormone replacement therapy use (female-no use, female-use both versus male), and eGFR.

For the visit 5 analysis to generate preliminary data: 4,246 participants are potentially eligible, among whom 588 go on to die of any cause by 2017. Of these 588, 81 die of a smoking related cancer (127 die of any cancer).

We do not expect that all proteins associated with the smoking variables are all biologically linked with carcinogenesis. Thus, in the amended application in which we will propose to use visit 2 and 3 proteins, we will investigate whether the smoking-associated proteins we find to be associated with cancer mediate the association between smoking and cancer. To do so, we will test the association between smoking status and smoking-related cancer mortality with and without smoking-associated proteins as a covariate. Then, for the mediating proteins identified in B), we will conduct pathway analysis to detect the biological pathways that may link the associations.

For the visit 2 analysis to generate preliminary data: 13,057 participants are potentially eligible. Among 1,555 participants who were subsequently diagnosed with a smoking-related cancer, 1,010 died of a smoking-related cancer.

Major challenges:

- Analysis A: Systematic differences between smokers and non-smokers beyond smoking that may be correlated with protein levels. We will adjust for covariates, but residual confounding could remain. In analysis we will do when visit 2 and 3 proteins are available, we will consider other covariates (e.g., related to SES) and strategies for deal with confounding.
- Analysis B: Small sample size for outcomes post visit 5. We will express proteins as continuous variables, which will increase power, but we are aware that this approach could mis-specify the shape of any association. When visit 2 and 3 proteins are available, we will consider other expressions of the proteins (e.g., threshold, quantiles).

7.a. Will the data be used for non-CVD analysis in this manuscript? ☒ 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? ____ 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 __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.csc.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)?

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* 2017.27 [proteomics], 2011.07 [cancer]; 1995.04 [cancer])

____ 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 ☒ No.

References

1. Warren GW, Alberg AJ, Kraft AS, Cummings KM. The 2014 Surgeon General's report: "The health consequences of smoking--50 years of progress": a paradigm shift in cancer care. *Cancer* 2014;120:1914-6.
2. Siegel RL, Jacobs EJ, Newton CC, et al. Deaths Due to Cigarette Smoking for 12 Smoking-Related Cancers in the United States. *Cigarette Smoking Deaths for Smoking-Related Cancers*. *Letters. JAMA Internal Medicine* 2015;175:1574-6.
3. Shiels MS, Pfeiffer RM, Hildesheim A, et al. Circulating inflammation markers and prospective risk for lung cancer. *Journal of the National Cancer Institute* 2013;105:1871-80.
4. Guida F, Sun N, Bantis LE, et al. Assessment of Lung Cancer Risk on the Basis of a Biomarker Panel of Circulating Proteins. *JAMA Oncol* 2018;4:e182078.
5. Ganz P, Heidecker B, Hveem K, et al. Development and Validation of a Protein-Based Risk Score for Cardiovascular Outcomes Among Patients With Stable Coronary Heart Disease. *JAMA* 2016;315:2532-41.
6. Ngo D, Sinha S, Shen D, et al. Aptamer-Based Proteomic Profiling Reveals Novel Candidate Biomarkers and Pathways in Cardiovascular Disease. *Circulation* 2016;134:270-85.
7. Williams SA, Murthy AC, DeLisle RK, et al. Improving Assessment of Drug Safety Through Proteomics: Early Detection and Mechanistic Characterization of the Unforeseen Harmful Effects of Torcetrapib. *Circulation* 2017;in press.