ARIC Manuscript Proposal #3499

PC Reviewed: 11/12/19  Status: _____  Priority: 2
SC Reviewed: _________  Status: _____  Priority: ____

1.a. **Full Title:** Prenatal exposure to smoking and adult cancer risk

b. **Abbreviated Title (Length 26 characters):** Prenatal smoking and cancer

2. **Writing Group:** Ladd-Acosta, Platz, Joshu, Fallin, Barber, Jones, and other interested ARIC investigators or members of the MD focused cancer priorities work group

I, the first author, confirm that all the coauthors have given their approval for this manuscript proposal. CL-A (ARIC author); EAP (ARIC author)

**First author:** Christine Ladd-Acosta, PhD
Address: Department of Epidemiology
Johns Hopkins Bloomberg School of Public Health
615 N. Wolfe Street, W6509
Baltimore, MD 21205
Phone: 443-287-5180
E-mail: claddac1@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 St., Room E6132
Baltimore, MD 21205
Phone: 410-614-9674
Fax: 410-614-2632
E-mail: eplatz1@jhu.edu

3. **Timeline:** We expect to complete this work by August 2020. We plan to use extant DNA methylation, cancer outcome, and self-reported smoking history data.

4. **Rationale:** Considerable strides have been made in reducing cancer incidence rates through identification of modifiable exposures (e.g. smoking, HPV). However, the risks identified to date do not fully explain cancer incidence rates. There is a need to identify additional cancer risk factors to determine their contribution and to better inform future prevention and/or intervention efforts. Previous studies to identify cancer risk factors have mainly focused on exposures during the adult time window but emerging evidence suggests prenatal and early life exposures influence adult health, including cancer.1-3 Several lines of evidence implicate the prenatal window as an important risk period for adult cancer risk.4-17
Direct and indirect evidence suggests prenatal exposure to smoking, in particular, may influence cancer risk in adulthood. Cigarettes contain toxicants that cross the placental barrier and have genotoxic effects on fetal cells.\textsuperscript{18-20} Thus, it is reasonable to propose maternal smoking may impact fetal processes and are related to their cancer development as adults. Other supportive indirect evidence includes parallel lines of investigation that have linked (a) prenatal smoking exposure to birth outcomes, and then (b) birth outcomes to adult cancer risk. For example, maternal smoking during pregnancy is strongly associated with decreased infant birth weight\textsuperscript{21} and decreased birth weight has been associated with increased ovarian\textsuperscript{22} and breast\textsuperscript{23} cancers. In addition, blood DNA methylation alterations detected in adult blood are associated with prenatal smoking, independent of own personal smoking history,\textsuperscript{24} and altered DNA methylation is critical to adult carcinogenesis.\textsuperscript{25} Direct investigation of the impact of prenatal smoking exposure on cancer risk has included research in animal models as well as humans. Mouse pups that were prenatally exposed to cigarette smoke showed higher susceptibility to cancer as adults.\textsuperscript{26} Human observational studies have shown associations between prenatal exposure to smoking and increased risk of nervous system and ocular tumors (e.g. neuroblastoma, retinoblastoma) but not leukemias or lymphomas in children.\textsuperscript{27-38} More limited research in adults has linked prenatal smoking exposure to increased colon cancer risk \textsuperscript{39} and to breast cancer risk.\textsuperscript{40-44}

While generally supportive of our hypothesis, there were considerable limitations to these previous studies including: (a) small sample sizes, (b) lack of prospective collection of prenatal exposure and subsequent cancer outcome data, and/or (c) failure to consider adult-onset or multiple types of adult cancers. Studies that overcome these issues are needed to rigorously evaluate the effects of prenatal smoking exposure on adult cancer risk. Ideally, this hypothesis would be tested using large cohorts with prospective measures of prenatal and adult smoking history, potential confounders, and unified adult cancer outcomes. However, prospective cohort study designs of this nature, i.e. those that seek to collect data across the life course, are particularly challenging due to their high cost and potential lower efficiency. These practical issues have undoubtedly contributed to the paucity of prospective evaluations of prenatal smoking exposure effects on adult cancer outcomes and are likely the reason for the larger body of evidence in childhood cancers, i.e. they require following individuals over a shorter period of time. ARIC is a long standing large cohort with rigorous prospective collection of multiple types of incident adult cancer, adult smoking history, and other potential covariates of interest. However, like most adult cohort studies, it did not measure prenatal smoking exposure. In theory, we could develop a questionnaire to collect maternal prenatal smoking history from ARIC participants but like previous studies it would be subject to recall bias, may have a high rate of exposure misclassification due to length of time between exposure and collection of data and/or lack of knowledge of mother’s pregnancy smoking history, and there would be additional study/participant burdens to obtain this information. In this proposal, we seek to overcome these study design limitations by measuring prenatal smoking exposure status using an adult blood DNA methylation biomarker. This enables us to obtain a measure of prenatal smoking exposure among participants in ARIC in an unbiased, robust, and cost-efficient way. Immediately below we summarize the evidence supporting the robustness and accuracy of this DNA methylation-based measure of prenatal smoking exposure, using adult blood DNA methylation measurements.
Large-scale studies have demonstrated that blood DNA methylation levels, at specific locations in the genome, are associated with prenatal exposure to smoking\textsuperscript{45}. The set of smoking-related methylation loci are often referred to as a “DNA methylation signature of prenatal smoking”. Distinct DNA methylation signatures of smoking have been found for different life stages of exposure including the prenatal\textsuperscript{24}, adolescent\textsuperscript{46}, and adulthood\textsuperscript{24} windows. Prenatal smoking methylation signatures have proven robust with replication in many thousands of samples from around the world\textsuperscript{45} and have been shown to persist into late adulthood\textsuperscript{24,47,48}. Predictive machine learning models\textsuperscript{49} and methylation scores\textsuperscript{50}, using adult blood DNA methylation measures, are highly predictive of prenatal exposure to maternal smoking and are independent of personal smoking history in the postnatal window\textsuperscript{24}. We plan to use extant adult blood DNA methylation data in ARIC and these robust methylation scores/predictive measures of prenatal smoking exposure, to evaluate prenatal smoking exposure associations with adult cancer risk.

5. Main Hypothesis/Study Questions:

Changes in fetal programming can have lifelong shifts in physiologic processes, organ structure/function, and response to the postnatal environment. Toxicants in cigarette smoke are known to cross the placental barrier and directly impact the genome of the developing fetus. \textbf{We hypothesize that prenatal smoking exposure is a risk factor for cancer incidence later in life, independent of personal smoking history}. We will test this hypothesis via two specific questions:

\textbf{Question 1: Does maternal smoking during pregnancy increase her offspring’s risk of having cancer as an adult?} We will test a model that includes a term for self-reported adult smoking history and prenatal smoking exposure (as measured via DNA methylation) to determine whether prenatal exposure to cigarette smoke is associated with incident cancer, after adjustment for adult smoking history (assessed via self-report and adult smoking methylation scores).

\textbf{Question 2: Does prenatal smoking exposure modify the effects of adult smoking history on incident cancer?} We will test a model that includes an interaction term for adult smoking history and prenatal smoking exposure to assess whether adult smoking history associations with cancer (assessed via self-report and adult smoking methylation scores) differ by prenatal exposure to smoking.

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).

\textbf{Study design:} Prospective cohort. Note: new measures of prenatal smoking exposure will be obtained using existing ARIC DNA methylation data and predictive machine learning models/scores.\textsuperscript{24,49,50}
Analytic sample: Our analytic sample will include all men and women with existing DNA methylation data from Visit 2 or 3 who consented to studies on chronic diseases including cancer as well as to genetic studies of this type (full). We will exclude any individuals with prevalent cancer, i.e. a cancer diagnosis at or before ARIC enrollment. The number of participants that meet our inclusion/exclusion criteria (n=3,625), and breakdown by incident cancer type, are summarized in the table immediately below. Note, unlike the overall ARIC cohort, the subset of ARIC participants with extant methylation data is predominantly from the Jacksonville MS site and Black (n=2,645).

Number of cancer cases overall and by type among eligible participants (N=3,625), ARIC Cancer 1986-2015

<table>
<thead>
<tr>
<th>Cancer Type</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Any</td>
<td>1,052</td>
</tr>
<tr>
<td>Sufficient sample size</td>
<td>657</td>
</tr>
<tr>
<td>Breast</td>
<td>156</td>
</tr>
<tr>
<td>Prostate</td>
<td>238</td>
</tr>
<tr>
<td>Hematopoietic</td>
<td>90</td>
</tr>
<tr>
<td>Colon</td>
<td>97</td>
</tr>
<tr>
<td>Kidney</td>
<td>42</td>
</tr>
<tr>
<td>Pancreas</td>
<td>34</td>
</tr>
<tr>
<td>Active smoking-related</td>
<td>171</td>
</tr>
<tr>
<td>Lung</td>
<td>124</td>
</tr>
<tr>
<td>Bladder</td>
<td>34</td>
</tr>
<tr>
<td>Esophageal</td>
<td>5</td>
</tr>
<tr>
<td>Head/neck</td>
<td>8</td>
</tr>
</tbody>
</table>

*Smoking is not a strong risk factor for incidence of these cancers, and thus we do not expect that smoking history will be associated with risk of these cancers. It is possible however, given previous conflicting evidence, that there is a slight elevated risk for breast cancer among prenatally exposed individuals.

Exposures:

Prenatal Smoking. This information will be obtained using an adult blood based DNA methylation signature of prenatal smoking exposure. DNA methylation measurements in ARIC were obtained using biospecimens collected at visit 2 or 3. We will use these extant ARIC DNA methylation data and previously identified prenatal smoking-methylation loci (n=19), from the largest EWAS studies to date, to compute a quantitative prenatal smoking exposure score. Higher methylation scores indicate the ARIC participants mother smoked during pregnancy; these scores robustly predict prenatal exposure, independent of postnatal and own adult smoking history. For each ARIC sample, we will compute a prenatal smoking methylation score as follows:
\[ \text{Methylation Score} = \sum_{i=1}^{n} \beta_i \text{CpG}_i \]

where:
- \( i \) denotes one locus, i.e. CpG site
- \( n \) denotes the total number of prenatal smoking associated loci to be summed

Note, we will use \( n = 19 \) loci that strongly predict prenatal smoking exposure from DNA methylation measures obtained in older adults\(^{24}\).

\( \text{CpG} \) denotes a normalized ARIC methylation value, computed as follows:

For each CpG with increased methylation related to prenatal smoking\(^{45}\), we will subtract the median methylation level reported among unexposed \(^{45}\) from the ARIC methylation level. For each CpG with decreased methylation related to smoking\(^{45}\), we will subtract the ARIC methylation level from the median methylation level reported among unexposed \(^{45}\).

\( \beta \) denotes a weighted effect size

this weighted effect size will be computed by dividing the methylation effect size reported in\(^ {45}\) by the average effect size of all measured CpGs in\(^ {45}\).

Adult smoking history will be obtained in 2 ways:

1. **Self-reported adult smoking history**: We will use self-reported smoking history data from visit 1 and visit 2 or 3, to correspond to the timing of when DNA methylation measurements were obtained, and define (a) categorical and (b) cumulative measures of smoking exposure. For categorical self-report exposure measures, individuals will be classified as former, never, or current smokers. For a cumulative measure of smoking exposure, we will use pack-years of cigarette smoking defined by multiplying the number of cigarette packs smoked per day by the number of years smoked. We will also use date started smoking and time-since quitting as a continuous measure of self-reported smoking history in our models. In the event that self-report smoking information is not available from visit 2 or 3 we will consider self-reported smoking data collected at baseline (visit 1).

2. **DNA methylation based adult smoking exposure (To be used for sensitivity analysis)**. DNA methylation measurements in ARIC were obtained using biospecimens collected at visit 2 or 3. We will use these extant ARIC DNA methylation data and previously identified adult smoking-methylation loci, from the largest EWAS to date\(^ {51}\), to derive quantitative methylation smoking scores that differ between: (1) current versus never smokers, (2) current vs. former vs. never smokers, as well as (3) a methylation score that captures cumulative lifetime dose of smoking, even after cessation for decades (similar conceptually to pack years). For each ARIC sample, we will compute a methylation score of adult smoking history as follows:
\[ Methylation\ Score = \sum_{i=1}^{n} \beta_i \text{CpG}_i \]

where:

- \( i \) denotes one locus, i.e. CpG site
- \( n \) denotes the total number of smoking associated loci to be summed
  - for Methylation Score (1), we will use \( n = 2,623 \) loci that differ in current and never smokers\(^{51}\)
  - for Methylation Score (2), we will use \( n = 185 \) loci that differentiate current, former, and never smokers\(^{51}\)
  - for Methylation Score (3), we will use \( n = 1,804 \) loci previously associated with number of pack years smoked\(^{51}\)

\( \text{CpG} \) denotes a normalized ARIC methylation value, computed as follows:
- For each CpG with increased methylation related to smoking\(^{51}\), we will subtract the median methylation level reported among nonsmokers in \(^51\) from the ARIC methylation level. For each CpG with decreased methylation related to smoking\(^{51}\), we will subtract the ARIC methylation level from the median methylation level reported among nonsmokers in \(^51\).

\( \beta \) denotes a weighted effect size
- this weighted effect size will be computed by dividing the methylation effect size reported in \(^51\) by the average effect size of all measured CpGs in \(^51\).

Note, we will independently test each of these adult smoking history measures in our analyses, as described in detail below.

**Outcomes:** Our decision to examine the broad outcomes below, as opposed to strongly smoking versus non-smoking was based on the rationale that the route of prenatal smoking, second-hand passive, is considerably different than for active first-hand smoking. Thus, the primary target tissues for cancer risk may differ. We therefore chose the outcomes below based on sufficient sample size and broad groupings.

1. **All cancers**
   - We will examine risk of any cancer, given that smoking is weakly to moderately a cause of cancer, beyond aerodigestive types. In addition, because the route of exposure under investigation here is trans-placental, exposure could impact multiple cell lineages during fetal development.

2. **Cancers with sufficient sample size:**
   - Prostate
   - Breast
   - Hematopoietic
   - Colon
We will study these cancers separately.

3. Active smoking-associated cancers:
   - Lung and bronchial
   - Esophageal
   - Bladder
   - Head/neck

We will study these cancers in aggregate as well as lung cancer separately.

**Data analysis.** We plan to use extant quality control filtered and normalized DNA methylation data for these analyses. Our data analysis plan involves 3 parts. First, we will compute DNA methylation scores of prenatal smoking exposures, as detailed above (in the “Exposures” section). Second, we plan to perform descriptive analyses and generate tables and plots showing the distribution of prenatal smoking DNA methylation scores in ARIC participants. In addition, we will determine the distribution of prenatal exposure scores by personal adult smoking history and other demographic factors such as race, ethnicity, sex, recruitment site, study visit, and methylation technical processing variables. The final step of our analysis will involve performing statistical analyses to address each of our 2 questions, as follows:

**Statistical analyses for Question 1:** Does maternal smoking during pregnancy increase her offspring's risk of having cancer as an adult?

We plan to perform time to event (incident cancer) analyses for each of the cancer outcomes detailed above. We will fit Cox proportional hazards regression models with incident cancer as our dependent variable and the methylation score as our independent predictor variable, with adjustment for personal history of adult smoking, age, race, and sex. Methylation scores will be modeled as a continuous variable, with hazard ratios (HR) calculated for a change in DNA methylation by 1 standard deviation. We will perform sensitivity analyses for alternative methylation measures of adult personal adult smoking history (details above, including self-report and adult smoking methylation scores). We will also consider inclusion of DNA methylation batch (which also corresponds to ARIC site), and blood cell composition adjustment variables in further models if our descriptive analysis suggests they may be related to cancer outcomes and methylation scores. We do not expect this to be a major problem because the extant DNA methylation data we will be using should already be adjusted for batch and the smoking related methylation loci used here have been shown previously not to be impacted by blood cell composition51. We will apply a p-value threshold of 0.05. Should we identify significant associations, we plan to also conduct ROC analysis with AUC comparisons for prenatal smoking methylation risk scores on their own and with adult smoking history.

**Statistical analyses for Question 2:** Does prenatal smoking exposure modify the effects of adult smoking history on incident cancer?
Similar to the statistical methods applied to address Question 1, for each cancer outcome indicated above, we will fit Cox proportional hazards regression models with incident cancer as our dependent outcome variable and the prenatal smoking methylation score as our independent predictor variable while controlling for age, race/ethnicity, and sex. Our model will also include an interaction term for adult smoking history and prenatal smoking exposure to assess whether adult smoking history associations with cancer (assessed via self-report and adult smoking methylation scores, independently) differ by prenatal exposure to smoking. Self-reported never/former/current smoking will be modeled as a categorical variable with never smoker as the reference group. Sensitivity analyses will also be performed using methylation scores for adult smoking which will be modeled as continuous variables, with hazard ratios (HR) calculated for a change of 1 standard deviation. We will also consider additional statistical models based on our descriptive statistics and results from question 1. We will apply a p-value threshold of $\leq 0.05$ for significant findings.

**Power.** The Table below provides the detectable effect sizes, i.e. change in methylation score of prenatal smoking exposure, for our adjusted Cox proportional hazards regression analyses at 80% power and a total of 3,625 individuals. We computed power using a 0.05 alpha threshold and a range of prenatal smoking exposure frequencies from 5% to 30%, given reported prevalence rates among women of childbearing age from 1920-1945 (representing the birth years of our ARIC participants included in this study; also see Figure on the right). \(^{52-58}\) We assumed that 74%, 96%, and 97% of never smokers survived overall, aerodigestive, and lung cancers, respectively.

We are reasonably well-powered to detect effect sizes that are in line with previously reported findings on prenatal smoking exposure and pediatric\(^{27-38}\) or adult cancers\(^{39-44}\), as also detailed in the rationale section above.

<table>
<thead>
<tr>
<th>Cancer Type</th>
<th>$\alpha=0.05$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall</td>
<td>1.14 to 1.32</td>
</tr>
<tr>
<td>Any Aerodigestive</td>
<td>1.34 to 1.80</td>
</tr>
<tr>
<td>Lung</td>
<td>1.44 to 2.08</td>
</tr>
<tr>
<td>Breast</td>
<td>1.36 to 1.86</td>
</tr>
<tr>
<td>Prostate</td>
<td>1.30 to 1.70</td>
</tr>
<tr>
<td>Hematopoietic</td>
<td>1.55 to 2.39</td>
</tr>
</tbody>
</table>

HR, Hazard Ratio
Anticipated Limitations or Challenges

One challenge of this work is that the magnitude of change in DNA methylation score (1 SD) may be difficult to interpret with respect to number of cigarettes or pack years smoked and effect on risk. To address this point, we could consider converting methylation scores to a more interpretable metric of smoking. This could also be addressed in future work should we show here that DNA methylation based measures of prenatal smoking are associated with increased cancer incidence. Furthermore, it is highly plausible that the same number of cigarettes or pack years smoked in pregnancy may result in very different levels of methylation changes because of differences in smoking behavior, genetic variation related to activation and detoxification of cigarette smoke constituents and DNA repair, etc (see Figure immediately below). Thus, we would not expect DNA methylation levels to correspond perfectly to external self-report smoking history metrics. In fact, because DNA methylation changes may better reflect the internal dose, or biologic response, to smoking it may be more relevant to the health effects and a strength of this study.

Second, although we adjust for race in our analyses, it is possible that we may need to compute ancestry-specific normalized beta-values when we derive our methylation scores as opposed to adjust in our regression models. Should this arise as an issue in our initial descriptive/characterization analyses we can go back and revise the scores using ancestry specific reference data, to the extent possible. Because race also corresponds to batch of methylation measurement in ARIC we will also consider performing analyses within each batch/race group followed by meta-analysis using METAL.

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 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? _X__ 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”? _X___ 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

We did not identify any manuscript proposals on this topic. The idea has been discussed by the ARIC Cancer Working Group

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)?

Ancillary study proposal for smoking and methylation:
2016.14 Transcriptomic and Epigenetic Signatures of Tobacco Smoking

2017.17 - Epigenome-wide correlation of DNA methylation and network analysis of smoking

MS proposals using the methylation data, include:

#3005 - DNA methylation-based risk score and prediction of all-cause mortality in the Atherosclerosis Risk in Communities Study, which includes cancer mortality

# 3211 - Socioeconomic Adversity, Epigenetics and Measures of Obesity in the ARIC Study

# 3247 - DNA Methylation Signatures of Life’s Simple 7 in Middle Aged Adults: The Atherosclerosis Risk in Communities Study

# 3248 - Preliminary Analysis of Obesity-related DNA Methylation Markers and Risk of Post-menopausal Breast and Endometrial Cancer

#3414 – Assessment of smoking-related cancer risk using DNA methylation as a measure of adult smoking history

We will reach out to Weihua Guan (ancillary study 2017.17) , Ellen Demerath, Jim Pankow and Eric Boerwinkle for collaboration and recommendations for authorship as we did for #3248 and #3414

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* 2011.07 (ARIC 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.cscc.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.cscc.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

25. Flavahan WA, Gaskell E, Bernstein BE. Epigenetic plasticity and the hallmarks of cancer. Science 2017;357.