ARIC Manuscript Proposal #3414

PC Reviewed: 6/18/19      Status: _____     Priority: 2
SC Reviewed: _______      Status: _____     Priority: ____

1.a. **Full Title**: Assessment of smoking-related cancer risk using DNA methylation as a measure of adult smoking history

**b. Abbreviated Title (Length 26 characters)**: Smoking biomarker and cancer

2. **Writing Group**: Ladd-Acosta, Platz, Joshu, Fallin, Barber, Jones, Guan, Demerath, 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 February 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) and efforts to reduce exposure to those factors. Identification of other modifiable exposure risks could provide new targets for prevention efforts and, thus, hold promise for further reduction in cancer incidence rates. However, current study designs often lack robust measures of environmental exposures which limits our ability to fully investigate and identify novel risk factors. For example, retrospective collection of exposure
information is subject to recall misclassification or bias (if collected after diagnosis). Even prospective ascertainment of exposure information could miss relevant risk windows and is limited in the number and type of exposures that can be feasibly ascertained. In addition, the biologic impact of “external” exposure measurements can differ between individuals due to inter-individual differences in rates of enzymatic detoxification and metabolism. Thus, “external” exposure measures may not appropriately capture “internal biologic dose” of exposure which may be more strongly associated with disease risk. Repeated prospective (or retrospective) biosampling enables circulating measures of specific toxicants that cannot be ascertained through medical records, clinical evaluations, or questionnaires but still have several substantial limitations. First, they could miss relevant windows of exposure. Second, they may be inaccurate for exposures with short half-lives (e.g. cotinine). Third, they often do not capture information about cumulative lifetime exposure (e.g. metals). New biomarkers of environmental exposures that can overcome current limitations are needed to rigorously test and identify exposures that influence cancer risk.

DNA methylation, a type of epigenetic modification, holds considerable promise as a robust exposure biomarker that can overcome exposure measurement limitations detailed above. It has been shown to capture cumulative lifetime exposure, effective biologic dose, and specific windows of exposure (from the prenatal to adulthood), in an unbiased way. The most robust and well developed findings to date have been for smoking exposure but DNA methylation signatures for other environmental factors, ranging from socioeconomic status and childhood adversity to metals and air pollutants continue to emerge. This area of research is in its infancy and additional research is needed to develop predictive models and exposure methylation scores as well as to evaluate their utility for testing associations between an environmental factor and cancer risk. The main goal of this project is to provide proof-of-principle evidence to support the utility and validity of using DNA methylation values themselves as a measure of exposure for risk factor identification and also to determine whether a methylation-based measure of exposure, i.e. “internal dose”, provides additional risk information, independent of self-reported measures of exposure. To carry out this proof-of-principle work, we have chosen to examine the effect of DNA methylation signatures of smoking exposure on aerodigestive cancer risk (lung, larynx, oral cavity and throat, bladder, esophageal) given their known strong causal relationship and also because of the robustness of DNA methylation signatures of smoking and maturity of the field relative to other exposures. Immediately below, we summarize evidence showing the potential for DNA methylation to serve as a measure of smoking exposure.

Several large-scale studies have demonstrated that DNA methylation levels, at specific locations in the genome, are associated with current, former, and never personal smoking exposures. Additionally, DNA methylation levels at these specific genomic loci reflect number of pack years smoked and time since cessation. In fact, smoking-related DNA methylation levels in blood have been shown to persist for at least 30 years after an individual quit smoking. A smoking methylation score, i.e. a single value that captures smoking status, can be derived from the results of the largest personal smoking EWAS to date (n=15,907) and was recently shown to accurately predict (AUC = 0.88) personal smoking exposure. Based on the sites identified by this EWAS, DNA methylation scores of smoking were derived and shown to explain a substantial proportion of the phenotypic variance, with 60.9% of the phenotypic variance in smoking explained by the methylation score. DNA methylation scores have also been shown to...
be substantially more predictive than polygenic risk scores for a number of exposures, including smoking, alcohol use, and BMI, suggesting DNA methylation patterns capture a biologic state that is closer to a phenotype than DNA sequence variants. Distinct smoking-related DNA methylation patterns have been found for different life stages of exposure including the prenatal, adolescent, and adulthood windows. For example, predictive machine learning models and methylation scores are highly predictive of prenatal exposure to maternal smoking and are independent of personal smoking exposure in the postnatal window.

Despite the robust methylation changes associated with smoking, few studies have used smoking-related methylation values themselves to test for exposure associations with disease risk. One study examined a small subset of all smoking-related DNA methylation loci (n=3) and found they were highly predictive (AUC = 0.80) of lung cancer incidence among 600 European individuals. The authors also showed the predictive ability of methylation levels at these 3 loci was greater than that of self-reported smoking status. A second study showed smoking-related DNA methylation patterns, derived from buccal epithelium, can accurately discriminate cancer from normal tissue. Finally, using blood DNA methylation data from 2 smoking-related loci, Zhang et al showed that individuals with the lowest quartiles of methylation at both loci, had increased risk of all-cause (aHR=3.59), cardiovascular (aHR=7.41), and cancer mortality (aHR = 2.48). Associations between self-reported smoking status and cancer mortality were attenuated when adjusted for DNA methylation levels at these loci. Although these 3 studies included DNA methylation levels related to smoking exposure in their analyses, limitations in sample size, study design, or type of methylation locus tested preclude a full evaluation and understanding of whether DNA methylation scores, i.e. a summed, weighted, value of exposure, can be used as a measure of exposure. Furthermore, studies to evaluate whether including both a DNA methylation score, as a marker of “internal biologic exposure dose”, and self-report are more predictive of cancer incidence than either measure alone are needed. We will overcome previous limitations in this area of research by examining: (1) a large number of samples from a more diverse population (Black and White participants), (2) an accessible tissue that is practical to collect in epidemiology studies (blood leukocyte DNA), (3) a composite DNA methylation score of smoking exposure using results from the largest smoking EWAS to date, (4) a well-characterized prospective cohort with over 30 years of follow up time, and (5) examining individual and joint contributions of self-report and methylation based measures of exposure, to fill critical gaps in our knowledge on the utility of DNA methylation values as a measure of exposure.

We expect that the results of this project will advance our knowledge about the utility and validity of using DNA methylation values as a measure of exposure, including past exposures. This could open new avenues of research to identify novel risk factors for cancer that would otherwise be difficult or impossible to carry out due to limitations in current study designs with reliable ascertainment of certain exposures. Additionally, our results will provide information about whether a biologic marker of smoking exposure contains additional risk information that is not captured by self-reported or cotinine based measures of smoking exposure.

5. Main Hypothesis/Study Questions:
In this proposal, we will address 2 main questions:
Question 1. Can DNA methylation values themselves be used to identify exposure-disease associations? To answer this question, we will examine a well-known risk factor for aerodigestive cancers - smoking history – as an exemplary model. *We hypothesize that DNA methylation scores of smoking will show strong associations with aerodigestive cancer risk.*

Question 2. Does a DNA methylation score of smoking history explain more variance in aerodigestive cancer risk than “traditional” self-reported smoking history (i.e., current status, intensity, start date, and time since quitting)? We will test a model that includes both self-reported and DNA methylation measures of exposure to assess whether exposure-cancer outcome associations persist after adjustment for self-reported smoking history. *We hypothesize that DNA methylation based measures of smoking history better reflect a “biologic dose” of exposure, are more proximal to the phenotype, and thus, more of the variance in cancer outcomes will be explained by DNA methylation measures of smoking than self-reported measures of smoking history.*

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:** We will use the prospective cohort design. The goal of this work is to address gaps in our knowledge about the utility of DNA methylation to serve as a measure of exposure when testing for exposure-disease associations. Data from a subset of 3,625 ARIC participants for whom methylation profiling has already been done will be analyzed, as detailed below, to determine the association between DNA methylation based smoking measures and incident aerodigestive cancer.

**Analytic population:** Our analysis 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 who reported a cancer diagnosis at or before the time of DNA methylation measurement (at visit 2 or 3). The number of participants that meet our inclusion criteria 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.

<table>
<thead>
<tr>
<th>Cancer Type</th>
<th>Total (n=3,625)</th>
<th>White (N=980)</th>
<th>Black (N=2,645)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Any</td>
<td>939</td>
<td>269</td>
<td>670</td>
</tr>
<tr>
<td><strong>Any Aerodigestive</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lung</td>
<td>114</td>
<td>34</td>
<td>80</td>
</tr>
<tr>
<td>Head/neck</td>
<td>6</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>Bladder</td>
<td>31</td>
<td>16</td>
<td>15</td>
</tr>
<tr>
<td>Esophagus</td>
<td>5</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td><strong>“Negative control” cancers</strong></td>
<td>365</td>
<td><strong>86</strong></td>
<td><strong>279</strong></td>
</tr>
<tr>
<td>Prostate</td>
<td>225</td>
<td>54</td>
<td>171</td>
</tr>
</tbody>
</table>
Smoking is not a strong risk factor for incidence of these cancers, and thus we do not expect that DNA methylation based smoking history will be associated with risk of these cancers.

**Exposures:**

Smoking. The effects of this exposure on cancer risk will be defined in several ways to address the main goals of this proposal:

*Self-reported 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. We will use smoking data collected at visits 2 or 3, corresponding to which visit the DNA methylation measures were obtained. 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).

*DNA methylation based 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 smoking-methylation loci, from the largest EWAS to date\(^3\), 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), and (4) showing persistent changes after 30 years of smoking cessation. In addition, we will derive a categorical variable that classifies each sample as a current, former, or never smoker based on the methylation score; we will use a random forests method to empirically define a methylation score threshold for classification, as described previously\(^15\). We will independently test each methylation score as well as the methylation defined smoking category in our analyses, as described in detail below.

Methylation measures of smoking will be computed as follows:

For each ARIC sample, we will compute a methylation score 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\textsuperscript{3}.

for Methylation Score (2), we will use n=185 loci that differentiate current, former, and never smokers\textsuperscript{3}.

for Methylation Score (3), we will use n= 1,804 loci previously associated with number of pack years smoked\textsuperscript{3}.

for Methylation Score (4), we will use n= 36 loci with smoking-related changes that persisted for 30 years after smoking cessation\textsuperscript{3}.

\textit{CpG} denotes a normalized ARIC methylation value, computed as follows:

For each CpG with increased methylation related to smoking\textsuperscript{3}, we will subtract the median methylation level reported among nonsmokers in \textsuperscript{3} from the ARIC methylation level. For each CpG with decreased methylation related to smoking\textsuperscript{3}, we will subtract the ARIC methylation level from the median methylation level reported among nonsmokers in \textsuperscript{3}.

\(\beta\) denotes a weighted effect size

this weighted effect size will be computed by dividing the methylation effect size reported in \textsuperscript{3} by the average effect size of all measured CpGs in \textsuperscript{3}.

For each ARIC sample, we will categorize them as a “current”, “former”, or “never” smokers using Methylation Score 2 and a random forests method to empirically select a threshold to use for classification.

DNA methylation based marker of multi-exposures: “epigenetic clock/aging”

We will also apply the Horvath epigenetic clock algorithm\textsuperscript{16} to existing ARIC epigenetic data to predict biologic age. Accelerated epigenetic aging, relative to one’s chronologic age, has been associated with numerous adverse health outcomes, including cancer, and is thought to capture the cumulative biologic effects of multiple exposures across the lifespan.

The main goal of using biologic age in this proposal is not to test prospective associations between accelerated aging and cancer incidence, per se, but rather to determine the extent to which smoking-related methylation changes are captured by the epigenetic clock or represent an independent cancer incidence risk signal. As part of question 2, we will first determine the correlation between the predicted epigenetic age, using epigenetic clocks, and DNA methylation based exposures scores. If there is not a strong correlation between the 2 methylation measures, we will employ a statistical model that tests joint and independent effects of: (a) self-reported smoking, (b) DNA methylation measures/scores of smoking, specifically, and (c) a methylation-based measure of multiple cumulative lifetime exposures, potentially including smoking, on cancer incidence.

Outcomes: We will examine the following outcomes

1. **Strongly smoking-related cancers:**
   - Lung and bronchial
   - Esophageal
   - Head/Neck/oral cavity
   - Bladder
We will study these cancers in aggregate as well as lung cancer separately.

2. **Cancers for which smoking is not a strong risk factor** (serve as negative controls):
   - Prostate
   - Breast

We will study these sex-specific cancers separately.

3. **All cancers**
   We will also examine risk of any cancer and after excluding breast and prostate cancers, given that smoking is weakly to moderately a cause of cancer even beyond aerodigestive.

**Data analysis.** We plan to use extant cleaned, quality control filtered, normalized extant DNA methylation data for these analyses. Our data analysis plan involves 2 main parts. First, we plan to perform descriptive analyses and generate tables showing the number and proportion of individuals with self-reported smoking history and methylation based measures of exposure by demographic factors such as race, ethnicity, sex, recruitment site, study visit, and methylation technical processing variables. We will also compare our DNA methylation based smoking scores to self-reported smoking data (current, former, never, and pack-years) to validate that they are performing as expected and for further characterization/insights into how the 2 measures of exposure compare. Second, we will carry out statistical analyses to address each question:

**Statistical analyses for Question 1:** *Can DNA methylation values themselves be used to identify smoking as a risk factor for cancer?*

We plan to perform time to event (incident cancer) analyses. First, for each cancer outcome indicated above, we plan to generate Kaplan-Meier plots to assess time to cancer incidence by methylation derived never, former, and current smoking status groups. We will also do this for each of our 3 smoking methylation scores, grouping them by quartiles. Associations will be determined via log-rank-tests. Next, for each of our methylation smoking metrics and each cancer outcome, we will fit Cox proportional hazards regression models with incident cancer as our dependent variable and the methylation measure as our independent predictor variable, with adjustment for age, race/ethnicity, 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, and methylation based “never/current/former” classifications modeled as a categorical variable with never smoker as the reference group. We will also consider inclusion of DNA methylation batch 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 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 composition. To conservatively account for multiple testing, we will apply a Bonferroni adjusted p-value threshold of $< 0.0125$ for significant findings and p-values $< 0.05$ will be considered suggestive. Should we identify significant associations, we plan to also conduct ROC analysis with AUC comparisons for methylation risk scores on their own and with baseline self-reported smoking history.
Statistical analyses for Question 2: Does a DNA methylation score of smoking history explain additional variance in cancer diagnosis above and beyond self-reported smoking and other epigenetic measures of exposure?

Similar to the statistical methods applied to address Question 1, for each cancer outcome indicated above, we will first generate Kaplan-Meier (K-M) plots to assess time to cancer incidence by smoking status – using self-reported and methylation-based measures of never/former/current exposure categories. We will utilize log-rank-tests to compare self- and methylation-based K-M curves. Similarly, we will plot and compare self-reported pack-years of smoking and cumulative smoking methylation scores. Second, for each methylation score and each cancer outcome, we will fit Cox proportional hazards regression models with incident cancer as our dependent outcome variable and smoking methylation score as our independent predictor variable while controlling for self-reported smoking history, epigenetic age, race/ethnicity, and sex. If there is not a strong correlation between DNA methylation scores of smoking and epigenetic age we will run models with both DNA methylation smoking scores and epigenetic age. Including epigenetic age in our model will enable us to evaluate whether smoking history effects on cancer risk are fully captured using a more general methylation measure of multiple exposures across the lifetime or whether it capture independent risk effects. Our main goal isn’t to assess epigenetic age acceleration on incident cancer but rather to tease apart whether the smoking methylation effects on incident cancer risk are captured via the epigenetic age metric or are independent. Methylation scores, epigenetic age, and self-reported pack year measures of smoking will be modeled as continuous variables, with hazard ratios (HR) calculated for a change of 1 standard deviation. Self-reported never/former/current smoking will be modeled as a categorical variable with never smoker as the reference group. We will also consider additional statistical models based on our descriptive statistics and results from question 1. To conservatively account for multiple testing, we will apply a Bonferroni adjusted p-value threshold of < 0.0125 for significant findings and p-values < 0.05 will be considered suggestive.

**Power.**

For our Kaplan-Meier log-rank test of differences in curve analyses, the table below provides the detectable effect sizes at 80% power, assuming a p-value threshold of 0.05 and 3,625 total samples (dichotomized into 2,175 ever and 1,450 never smokers). We assumed that 74%, 96%, and 97% of never smokers survived overall, aerodigestive, and lung cancers, respectively. These effect sizes are an order of magnitude lower than what has been repeatedly observed for self-report smoking history associations with cancer; thus, we are well powered for this analysis. For our adjusted Cox proportional hazards regression analyses where DNA methylation scores are continuous quantitative measures we expect similar, if not improved, power.

<table>
<thead>
<tr>
<th>Minimum detectable effect sizes at 80% power:</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cancer Type</strong></td>
</tr>
<tr>
<td>Overall</td>
</tr>
<tr>
<td>Any Aerodigestive</td>
</tr>
<tr>
<td>Lung</td>
</tr>
<tr>
<td>Breast</td>
</tr>
<tr>
<td>Prostate</td>
</tr>
</tbody>
</table>
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 smoking can be used to detect disease associations. Second, we cannot directly compare self-reported pack-years to methylation based cumulative exposures due to differences in the unit of measurement, so have limited our direct comparisons to smoking classification categories (current/former/never). Third, although we adjust for race in our analyses and African American and Caucasian individuals were used for select the loci we test, 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 as ancestry specific reference data exists. Because race also corresponds to batch of methylation measurement in ARIC we will also consider performing analyses within each batch/race and then meta-analyze the results using METAL. Finally, for methylation score 1, we use CpG sites that were initially selected to differentiate current from never smokers. We wanted to include this measure for 2 reasons: (a) it should capture strong differences in smoking-related cancer incidence risk, and (b) we can use it to validate our score generation methods because it has been used and well-validated in the past by other groups. However, it is not clear where the former smokers will lie in the score distribution; if that does not hold in our initial descriptive statistic evaluation, we could, as an alternative, generate methylation score 1 among ARIC individuals that reported current or never smoking status only.

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
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

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.

11. 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.csc.c.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.csec.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