ARIC Manuscript Proposal #1965

PC Reviewed: 7/10/12	**Status: A**	**Priority: 2**
SC Reviewed: _____	**Status: _____**	**Priority: ____**

1.a. Full Title: Development of a quantitative model for the relative risk of cardiovascular disease by pack-years of cigarette smoking and assessment of modification of the strength of association by smoking rate and other host characteristics in the Atherosclerosis Risk in Communities Study

b. Abbreviated Title (Length 48 characters): Quantitative modeling of CVD risk and cigarette smoking

2. Writing Group:
Jay H. Lubin, Hiroshi Yatsuya, Pamela L. Lutsey, Mark Woodward, Rachel R Huxley

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

First author: Jay H. Lubin
Address: Biostatistics Branch
Division of Cancer Epidemiology and Genetics
National Cancer Institute, National Institutes of Health
6120 Executive Blvd, Room 8105
Rockville, MD 20852
Phone: 301.496.3357 Fax: 301.402.0081
E-mail:lubinj@mail.nih.gov

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: Rachel Huxley
Address: Div of Epidemiology & Community Health
University of Minnesota
1300 S 2nd St, suite 300. Minneapolis, MN 55454
Phone 952-250-1730
E-mail: rhuxley@umn.edu

3. Timeline:
Data analysis, 4 months; first draft of the manuscript, 4 months

4. Rationale:
Cigarette smoking is an established risk factor for cardiovascular disease (CVD) (2, 12, 13). Unraveling the molecular basis of the smoking and CVD association, including mechanisms of action of various constituents of tobacco smoke and the genetic basis of smoking persistence, is the focus of current epidemiologic research (14). The precise

quantitative characterization of disease risk with total cigarette smoke exposure, smoking intensity and duration can aid in understanding the biological basis of host reaction to the many chemical constituents in tobacco smoke. Previous investigators have evaluated CVD risk and individual smoking variables; however, there has been no comprehensive quantitative analysis of the association of total pack-years, the impact of delivery rate on the relative risk (RR) and the manner by which other risk factors modify the association.

Epidemiologic analyses typically compute the joint RRs for smoking intensity, as measured by cigarettes per day (CPD), and duration relative to never-smokers, or estimate RRs by CPD adjusted for smoking duration (or RRs by smoking duration adjusted for CPD). Interpretation of such results however is problematic since RRs for CPD at a fixed duration embed the effect of increasing pack-years. For example, at 30 years smoking, comparisons of RRs at 20 CPD and 30 CPD necessarily reflect different total exposures, i.e., 30 and 45 pack-years, respectively. Thus, RRs by CPD and duration cannot be interpreted as separate "independent" effects. In contrast, we propose to model total exposure (pack-years) and exposure rate (CPD), which reformulates analysis in terms of RR patterns with total exposure and modification of those patterns by delivery rate, where delivery rate represents the relative consequences of higher exposure rates for shorter durations compared with lower exposure rates for longer durations for a constant pack-years. The assessment of delivery rate is of particular interest, as it describes whether individuals with the same total exposure incur less, equal or more disease risk as exposures are protracted (lower rates for longer durations). This framework has yielded relatively simple characterizations of joint associations and has generated novel etiologic insights for various cancer outcomes and environmental exposures (cigarette smoking, alcohol consumption and inhaled arsenic containing dusts) (7-9).

In addition to smoking, studies link CVD risk with several factors, such as sex, race, physical activity, body mass index (BMI), alcohol consumption, age and various co-morbidities (total cholesterol, blood pressure [BP], type 2 diabetes, etc.) (12). A precise quantitative model for smoking-related CVD risk enables a more specific examination of effect modification, whereby factors may modify risk by influencing RR trends with pack-years or effects of delivery rate or both.

Studies of exposure to second hand smoke (SHS) support a causal link with CVD; however, substantial uncertainty remains regarding its magnitude of risk (10, 13). Exposure to SHS may correspond roughly to an estimated exposure to 0.5 to 1 CPD relative to active smoking, while risks are about 25%-30% above those never-exposed which is larger than expected based on RRs among active smokers (13). This suggests a non-linear relationship (6). Among ARIC participants, estimates of SHS exposure derive from the question, "During the past year, about how many hours per week, on the average, were you in close contact with people when they were smoking? For example, in your home, in a car, at work or other close quarters?" Since SHS information reflects current status only, we propose to examine whether current SHS exposure (ever/never and hours/wk) modifies the quantitative smoking-related model for CVD risk in active smokers.

With its large size and extended follow-up, the ARIC study represents an ideal opportunity to study the quantitative association of smoking, the impact of delivery rate and the role of effect modifiers on the smoking relationship. Results will aid the interpretation of smoking-related molecular and genetic factors and the development of improved risk prediction.

5. Main Hypothesis/Study Questions:

We propose to develop a quantitative RR model for CVD by total pack-years of cigarette smoking while accounting for the delivery rate of smoking and to assess various risk factors as effect modifiers. This proposal has three aims:

(i) to develop a statistical model that characterizes the quantitative relationship of the RR of CVD in terms of pack-years of cigarette smoking and the influence of delivery rate (the relative impact of simultaneously increasing CPD and decreasing duration of smoking for total pack-years held constant);

(ii) to evaluate potential effect modifiers of smoking-related RR patterns by various host characteristics, including sex, race, education, income, sports index, BMI, alcohol consumption, attained age, age at smoking initiation, years since smoking cessation and measures of co-morbidity (total cholesterol, LDL-C, BP and BP medications, type 2 diabetes, etc.); and

(iii) to evaluate the metrics of current SHS exposure as modifiers of smoking-related RR patterns of CVD risk.

Results from the recent analysis of Huxley and colleagues support a role for CPD as a modifier of the strength of the association between pack-years and CVD, and suggest that the pack-years-related CVD risk depend on smoking intensity {Huxley, 2012 831 /id}.

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).

We will assess the association of smoking and smoking cessation with morbidity and mortality from CVD using an approach appropriate to cohort data.

Primary exposure

Smoking information was measured during four study visits and at yearly intervals via a telephone questionnaire. We will exploit these data by incorporating time-independent (baseline) and time-dependent (longitudinal) information on smoking status, intensity, duration and cessation.

Outcomes of interest

We will include incident cases and deaths through the end of 2011 from CVD identified using ARIC adjudicated events for CHD, stroke (total, ischemic and hemorrhagic), heart

failure, total CVD, and CV mortality, and discharge codes from hospitalizations and death certificates for other outcomes. Events will be ascertained by annual follow-up interview and surveillance of hospital discharges in the ARIC study areas. Events will be validated by abstraction of hospital discharge records and death certificates, followed by classification according to ARIC study criteria including trained physician reviewers. Out-of-hospital deaths will be ascertained through death certificates and, when available, coroner or autopsy reports. CHD will be defined as a validated or definite or probable hospitalized MI, a definite CHD death, an unrecognized MI defined by electrocardiographic reading or coronary revascularization. Stroke events will be defined as a validated definite or probable hospitalized ischemic (or hemorrhagic) stroke confirmed by imaging.

Exclusions
Study participants will be excluded from the analysis if they fulfill at least one of the following criteria:

- missing data on smoking status at any of the study visits for which they were able to attend (unless status can be reasonably inferred from available information)
- prevalent cardiovascular disease, respiratory disease or cancer at study baseline
- ethnicity other than Black or White

Data structure
We assume a standard format for the data which is consonant with a prospective time-to-event cohort study that includes both time-independent and, to the extent possible, time-dependent variables, for use in a Cox proportional hazards regression model. We will use age as the time variable, allowing for age at cohort entry to define start of follow-up. However, the models described below are equally applicable for cohort data summarized in a multi-dimensional contingency table of disease counts and person-years and analyzed using Poisson regression, or for a sex and age-matched nested case-control sample analyzed using conditional likelihood regression. Due to the large number of disease events and the presence of multiple time-dependent covariates, we may ease the computation burden by using either a Poisson regression or a nested case-control approach. We will make a final decision during preliminary analyses of data.

Statistical analysis
Aim (i): We will model $r(t, z, p, n)$, the rate of CVD at age t for vector of adjustment covariates z, pack-years p and cigarettes per day n, as $r(t, z, p, n) = r_o(t)\, RR(z, p, n)$, where $r_o(t)$ is the baseline hazard function,

$$RR(z, p, n) = \exp(\alpha z)\, RR(p, n) \qquad (1)$$

and α is a vector of parameters associated with the adjustment variables.

We will first compute RRs for the cross-classification of pack-years and CPD categories relative to never-smokers, and plot RRs by pack-years categories within each of J CPD

categories. We will examine linearity of the (continuous) pack-years association using the model:

$$RR(p, n) = 1 + \sum_j \gamma_j p_j \quad (2)$$

where γ_j is the ERR/pack-year, which represents the strength of the association, and $p_j = p$ for CPD within the j^{th} category and zero otherwise. We will replace p_j with $p_j \times \exp(\theta_j p_j)$ to test the null hypothesis of no departures from linearity (i.e., $\theta_j=0$). Previous analyses suggest that linearity within CPD categories represents a useful first-order approximation to RR patterns. If so, we will fit a model of the form:

$$RR(p, n) = 1 + \beta\, p\, g(n) \quad (3)$$

where g(.) is a function that describes the modification of the strength of association (β) with continuous CPD, i.e., the delivery rate effect. Previous analyses have used $g(n) = \exp\{\varphi_1 \ln(n) + \varphi_2 \ln(n)^2\}$; however, we will evaluate alternatives, including restricted cubic splines (3, 5). We will verify the approach by comparing a plot of the estimates $\gamma_1,\ldots,\gamma_J$, by category-specific mean CPDs with the fitted $\beta\, g(n)$. Decisions will be aided by the Akaike Information Criterion (AIC) (1).

Since CVD risk is reduced with smoking cessation, we will fit the models to all data and to the subset of never, current and recent (<2 years) former smokers.

Aim (ii): We will consider effect modification by a categorical factor (*f*) with S levels using

$$RR(p,n,f) = 1 + \sum_s \beta_s p_s g_s(n_s) \quad (4)$$

where distinct β_s parameters and $g_s(.)$ functions replace β and g(.), and where p_s equals *p* and n_s equaled *n* within level s and zero otherwise. We will use deviances to compare model fit and evaluate whether effect modification derives from pack-years (different β's), delivery rate (different g(.) functions) or both. Starting with model (4), we will constrain the β's and/or g(.) functions to be equivalent across *f* and examine degradation in model fit. This approach, in contrast to starting with model (3) and enlarging the model, allows the evaluation of the interaction of *f* and one factor (e.g., pack-years) while minimizing influence of the interaction of *f* and its closely related correlate (e.g., CPD).

We will examine various potential effect modifiers of the smoking relationship, including attained age, sex, race, education, income, alcohol intake, sport activity, BMI, systolic BP, antihypertensive medications, type 2 diabetes, total cholesterol, use of filter/non-filter cigarettes, age at smoking initiation and years since cessation of smoking.

Aim (iii): We will evaluate SHS as a modifier of the smoking-related CVD risk using model (4) as described in (ii). Since never-smokers may have current SHS exposure, we will include the SHS metric as a risk/adjustment factor in the *z* vector to obtain an estimate of the main effects of SHS in never-smokers.

We will fit all models, conduct likelihood ratio tests and compute 95% confidence intervals (CI) (Wald-based for estimates of φ and γ and likelihood-based for estimates of β) using the Epicure suite of programs (11).

Adjustment variables
Where possible, we will adjust all analyses for age, sex, race, study site, education, income, usual alcohol intake, use of other tobacco products (pipes, cigars, snuff and chewing tobacco), sports activity, BMI, systolic BP, antihypertensive medication, type 2 diabetes and total cholesterol.

Limitations
Our analyses rely on self-reported smoking habits. While smoking status is generally reliably reported, there is potential misclassification in smoking amount and duration, although misclassification cannot be differential. The main concern is that misclassification may incorrectly classify some individuals as never-smokers or former-smokers and thus dilute the magnitude of the observed associations.

While patterns and effect modifications of the ERR/pack-year may reflect biologic phenomena, observed patterns may also reflect influences of nicotine satiation, whereby the disease-related yield per cigarette decreases with increasing intensity as smokers seek only to maintain addiction-sufficient nicotine levels, such that the number of cigarettes/day increasingly overestimates the true internal exposure rate. However, an analysis in a large lung cancer study found no evidence of a relationship between frequency or depth of inhalation and intensity, after controlling for total pack-years. Further, a sensitivity analysis, which estimated the degree of overestimation of internal exposure rate by CPD using cotinine as a marker of internal exposure rate, found that conclusions regarding delivery rate effects were largely unchanged, although power to evaluate the effects declined. Depending on results for CVD, we will conduct a sensitivity analysis to evaluate potential overestimation of internal exposure.

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 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

8.c. If yes, is the author aware that the participants with RES_DNA = 'not for profit' restriction must be excluded if the data are used by a for profit group?
____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)?

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* 2008.09)
___ 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/

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

1. Akaike H Information theory and an extension of the maximum likelihood principal. In: Petrov EB, Csaki F, eds. 2nd Annual Symposium on Information Theory and Control. Budapest, Hungary. Akademia Kiado, 1973:267-81.

2. Burns DM. Epidemiology of smoking-induced cardiovascular disease. Prog Cardiovasc Dis 2003;46:11-29.

3. Durrleman S, Simon R. Flexible regression-models with cubic-splines. Stat Med 1989;8:551-61.

4. Huxley RR, Yatsuya H, Lutsey PL, et al. Impact of age at smoking initiation, dosage, and time since quitting on cardiovascular disease in African Americans and Whites. Am J Epidemiol 2012;175:816-26.

5. Korn EL, Graubard BI. Analysis of Health Surveys. New York, NY: John Wiley and Sons, Inc., 1999,

6. Law MR, Wald NJ. Environmental tobacco smoke and ischemic heart disease. Prog Cardiovasc Dis 2003;46:31-8.

7. Lubin JH, Alavanja MCR, Caporaso N, et al. Cigarette smoking and cancer: modeling total exposure and intensity. Am J Epidemiol 2007;166:479-89.

8. Lubin JH, Purdue M, Kelsey KT, et al. Total exposure and exposure rate effects for alcohol and smoking and risk of head and neck cancer: a pooled analysis of case-control studies. Am J Epidemiol 2009;170:937-47.

9. Lubin JH, Moore LE, Fraumeni JF, et al. Respiratory cancer and inhaled inorganic arsenic in copper smelters workers: a linear relationship with cumulative exposure that increases with concentration. Environ Health Perspect 2008;116:1661-5.

10. National Cancer Institue Smoking and Tobacco Control Monograph 10: Health Effects of Exposure to Environmental Tobacco Smoke. Bethesda, MD: U.S. Department of Health and Human Services, Public Health Service, National Institutes of Health, National Cancer Institute. NIH Publication No. 99-4645, 1999.

11. Preston DL, Lubin JH, Pierce DA, et al. Epicure User's Guide. Seattle, Washington, USA: HiroSoft International Corporation, 2006,

12. U.S.Department of Health and Human Services The Health Consequences of Smoking: A Report of the Surgeon General. Washington D.C. 20402: U. S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health, Superintendent of Documents, U.S. Government Printing Office, 2004.

13. U.S.Department of Health and Human Services The Health Consequences of Involuntary Exposure to Tobacco Smoke: A Report of the Surgeon General. Washington D.C. 20402: U. S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health, Superintendent of Documents, U.S. Government Printing Office, 2006.

14. U.S.Department of Health and Human Services How Tobacco Smmoke Causes Disease: The Biology and Behavioral Basis for Smoking-Attributable Disease: A Report of the Surgeon General. Washington D.C. 20402: U. S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health, Superintendent of Documents, U.S. Government Printing Office, 2010.