ARIC Manuscript Proposal #3230

PC Reviewed: 9/11/18  Status: _____  Priority: 2
SC Reviewed: _________  Status: _____  Priority: _____

1.a. Full Title: Derivation of a COPD Risk Score

b. Abbreviated Title (Length 26 characters): Derivation of a COPD Risk Score

2. Writing Group:
   Writing group members:
   Surya P. Bhatt, University of Alabama at Birmingham, Birmingham, AL, USA, sbhatt@uabmc.edu
   Pallavi Balte, Columbia University, New York, NY, USA, ppb2119@cumc.columbia.edu
   Pat Cassano, Cornell University, Ithaca, NY, USA, pac6@cornell.edu
   David Couper, University of North Carolina, Chapel Hill, North Carolina, USA, david_couper@unc.edu
   David Jacobs, University of Minnesota, Minneapolis, MN, USA, jacob004@umn.edu
   Ravi Kalhan, Northwestern University, Chicago, IL, USA, RKalhan@nm.org
   Ana Navas Acien, Columbia University, New York, NY, USA, an2737@cumc.columbia.edu
   Robert Kaplan, Albert Einstein College of Medicine, New York, NY, USA, Robert.kaplan@einstein.yu.edu
   Anne Newman, University of Pittsburgh, Pittsburgh, PA, USA, NewmanA@edc.pitt.edu
   George O’Connor, Boston University, Boston, MA, USA, goconnor@bu.edu
   Joseph Schwartz, Columbia University, New York, NY, 10032, jes2226@cumc.columbia.edu
   Sachin Yende, University of Pittsburgh, Pittsburgh, PA, USA, Yendes@upmc.edu
   Jason L. Sanders, Massachusetts General Hospital, Boston, USA, Jsanders2@partners.org
   Paolo Chaves, Benjamin Leon Center for Geriatric Research and Educations, Paulo.Chaves@fiu.edu
   Wendy White, Tougaloo College, Tougaloo, MS, USA, wendywhite2001@yahoo.com
   Elizabeth C Oelsner, Columbia University, New York, NY, USA, eco7@cumc.columbia.edu
   TBD

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

First author:
Address:  THT 422, 1720 2nd Ave S, Birmingham, AL 35294

Phone:  205-934-5555  Fax:  205-934-3277
E-mail:  sbhatt@uabmc.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:  David Couper
Address:  137 East Franklin Street, Suite 203, CB #8030, Chapel Hill, NC 27599
3. **Timeline:** Manuscript submission by June 2019.

4. **Rationale:**
Chronic obstructive pulmonary disease (COPD) is the fourth leading cause of mortality in the United States and is associated with significant morbidity and healthcare costs.(1, 2) In contrast to cardiovascular disease (CVD) and cancer mortality, which have improved dramatically with widespread screening and promotion of primary preventative interventions, age-standardized COPD mortality has continued to rise. Hence, identification of persons with increased COPD risk and/or “early” COPD is urgently needed to target risk reduction strategies (e.g., smoking cessation) as well as to enroll high-risk individuals in primary prevention trials.

One particularly successful approach to CVD risk stratification and prevention has been the derivation, publication, and clinical application of cardiac risk scores such as the Framingham Risk Score and the more updated ASCVD Pooled Cohorts Risk Equation.(3, 4) The ASCVD risk equation was designed to leverage data that primary care providers could easily collect and that could be implemented in routine clinical practice. These data were used to estimate 10-year risk of occurrence of coronary death or fatal stroke, or first occurrence of nonfatal myocardial infarction or stroke. Derivation of this equation was accomplished using standard statistical methods in a pooled sample of four NHLBI Cohorts (ARIC, CHS, CARDIA, and Framingham Heart Study [Original and Offspring]).

Respiratory outcomes data from nine NIH-funded cohort studies, including the four cohorts used for the ASCVD Pooled Cohorts Risk Equations, have recently been harmonized and pooled to develop the NHLBI Pooled Cohorts Study. We therefore propose for the first time to derive, validate, and describe a COPD Pooled Cohorts Risk Equation using these data. We will specifically test the extent to which spirometry screening, which is not currently recommended in the US, could improve 10-year COPD risk prediction in smokers versus never-smokers.

5. **Main Hypothesis/Study Questions:**
- We will be able to derive a 10-year COPD Pooled Cohorts Risk Equation by applying similar methods as the ASCVD Pooled Cohorts Risk Equation, including restriction of predictors to major clinically available data (i.e., excluding spirometry)
- We will be able to describe temporal/cohort variations in absolute 10-year COPD risk
- We will be able to quantify the extent to which obtaining and modeling spirometry data improve 10-year risk estimation in smokers versus never-smokers
- We will be able to identify a threshold for 10-year COPD risk that is optimally predictive (best sensitivity/specificity) of incident COPD, as well as identifying thresholds for “high/pre-COPD” “intermediate,” and “low” risk.
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).

Sample

We propose to use nine cohorts with longitudinal spirometry, clinical, and events data that have been harmonized and pooled as part of the NHLBI Pooled Cohorts Study:
1. Atherosclerosis Risk in Communities (ARIC) Study
2. Coronary Artery Risk Development in Young Adults (CARDIA) Study
3. Cardiovascular Health Study (CHS)
4. Framingham Heart Study (FHS)
5. Health Aging and Body Composition (Health ABC) Study
6. Hispanic Community Health Study/Study of Latinos (HCHS/SOL)
7. Jackson Heart Study (JHS)
8. Multiethnic Study of Atherosclerosis (MESA)
9. Strong Heart Study (SHS)

Participants with prevalent COPD (defined by self-reported COPD) will be excluded. Follow-up will be truncated at 10 years, as in the derivation of the ASCVD Pooled Cohorts Risk Equation.

Most of the required data has already been harmonized and pooled at Columbia University, where the proposed analyses will be performed.

Endpoints:
• **Primary endpoint**: occurrence of COPD mortality or first occurrence of nonfatal COPD hospitalization (hospitalizations adjudicated or administratively coded as caused by COPD, chronic bronchitis, or emphysema [ICD-9 490-492, 496, 506.4; ICD-10 J40-J44]; for primary analyses, COPD may be classified as the primary/underlying or a contributing cause)
• **Secondary endpoints**
  o Occurrence of COPD mortality or first occurrence of nonfatal COPD hospitalization, requiring COPD to be classified as the primary/underlying cause (5)
  o Respiratory mortality
  o Incident airflow limitation (pre-bronchodilator FEV1/FVC < 0.70), defined among participants without baseline airflow limitation
• Competing risks: CVD mortality; cancer mortality; non-COPD respiratory mortality; non-CV and non-respiratory mortality

Risk factors:
• **A priori risk factors**
  ▪ Age
  ▪ Sex
  ▪ Race/ethnicity
  ▪ Smoking (never, former, current smoker, pack-years)
  ▪ Self-reported asthma
  ▪ Inhaler use
    ▪ Respiratory symptoms (Wheeze, Dyspnea, Bronchitis)
• Potential additional risk factors
  ▪ Educational attainment
  ▪ Height
  ▪ Weight
  ▪ Secondhand smoke exposure (hours/week)
  ▪ Cigarettes per day
  ▪ Years since cessation
  ▪ Age at smoking initiation
  ▪ Comorbidity (Coronary artery disease, Hypertension, Hyperlipidemia, Diabetes, eGFR, Albuminuria)

  o Lung function
    ▪ FEV1
    ▪ FVC
    ▪ FEV1/FVC
    ▪ FEF 25-75

Modeling strategy
• Competing risks models will be used including all *a priori* risk factors. Both age and age-squared will be modeled. Whether model fit is better using untransformed or natural log-transformed values for continuous predictors (e.g., age) will be evaluated. Interactions with age will be tested for each risk factor and retained in final models if the *p* value for the interaction term is less than .01, or the *p* value is .01 to .05 and the continuous net reclassification improvement for nonevents is 15 percent or greater, or the integrated discrimination improvement index (IDI) is statistically significant.(6, 7).
• Potential additional risk factors will be evaluated for potential improvement in model performance based on the framework of Hlatky, et al., 2009.(8) Improvement in discrimination will be defined as a relative IDI (rIDI) of 6 percent or more.
• Final model fit will be evaluated through the area under the receiver operating curve (*C*-statistic) for discrimination(9) and the calibration chi-squared statistic.(10) The importance of using race- and sex-specific models will be evaluated by comparative model fit.
• As noted above, follow-up for the derivation cohort will be truncated at 10 years, as in the derivation of the ASCVD Pooled Cohorts Risk Equation. Subsequent follow-up (post-10-year) will be used to create multiple validation cohorts for testing the derived score.
• Following the methods for derivation of 10-year hard ASCVD risk, the approach for estimating 10-year COPD risk will be as follows:
  o Risk factor values for a given individual are recorded/calculated (e.g., age, race/ethnicity, height, weight, smoking status, inhaler use, respiratory symptoms). Any appropriate interaction terms will then be calculated. These values will then be multiplied by the coefficients from the best model equation.
  o The sum of the “Coefficient×Value” will then calculated for the individual.
The estimated 10-year risk of COPD event will be formally calculated as 1 minus the survival rate at 10 years, raised to the power of the exponent of the “Coefficient×Value” sum minus the overall mean “Coefficient×Value” sum; or, in equation form:

\[1 - S_{10}^{e^{(\text{IndX}'B - \text{MeanX}'B)}}\]

In addition to the base model with the well-known risk factors and demographics listed above, we will also test addition of (where available) other parameters such as blood biomarkers, environmental pollution indices, area of residence, rural/urban residence, etc. We will retain these additional variables in the model if there is an improvement in discrimination of 5% or more.

**Pre-COPD**
We will assess the distribution of the 10-year COPD risk score and categorize risk into very high, high, intermediate, low and very low. The intermediate category will be deemed to have Pre-COPD.

**Limitations**
We acknowledge that respiratory hospitalization can be misclassified and also be biased by preexisting knowledge of diagnosis. However, unlike cardiac events, respiratory events are not characterized by elevated levels of any biomarkers and clinical diagnosis is the current gold standard.

**Summary/conclusion:** We anticipate the derivation of a robust Risk Score for COPD and the identification of a Pre-COPD stage that is characterized by an intermediate risk for poor respiratory outcomes. We anticipate that this Risk Score will be clinically applicable for the identification of subjects at risk for and with presence of COPD.

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? ____ 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)? None that are similar to the one proposed.

11.a. Is this manuscript proposal associated with any ARIC ancillary studies or use any ancillary study data? _____ Yes  ___X___ No

11.b. If yes, is the proposal

___ A. primarily the result of an ancillary study (list number* __________)

___ 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.
References