ARIC Manuscript Proposal #2863

PC Reviewed: 10/11/11     Status: _____     Priority: 2
SC Reviewed: _________     Status: _____     Priority: ____

1.a. Full Title: Clinical significance of lung function trajectory in six US population-based cohorts: the NHLBI Pooled Cohorts Study

b. Abbreviated Title (Length 26 characters): Lung function trajectories

2. Writing Group:

- First author: Elizabeth C Oelsner, Columbia University, New York, NY, USA, eco7@cumc.columbia.edu
- Co-authors:
  - 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
  - Paul Enright, University of Arizona, Tuscon, AZ, USA, lungguy@gmail.com
  - Aaron Folsom, University of Minnesota, Minneapolis, MN, USA, folso001@umn.edu
  - David Jacobs, University of Minnesota, Minneapolis, MN, USA, jacob004@umn.edu
  - Ravi Kalhan, Northwestern University, Chicago, IL, USA, RKalhan@nm.org
  - Richard Kronmal, University of Washington, Seattle, WA, USA, kronmal@u.washington.edu
  - Laura Loehr, University of North Carolina, Chapel Hill, NC, USA, lloehr@email.unc.edu
  - Stephanie London, NIH/NIEHS, Research Triangle park, NC, USA, london2@niehs.nih.gov
  - R Graham Barr, Columbia University, New York, NY, USA, rgb9@cumc.columbia.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
  - Lewis Smith, Northwestern University, Chicago, IL, USA, ljsmith@northwestern.edu
  - Wendy White, Tougaloo College, Tougaloo, MS, USA, wendywhite2001@yahoo.com
3. **Timeline:** Data from all cohorts have already been obtained, harmonized, and pooled under the auspices of approved paper/ancillary study proposals relating to specific biological hypotheses that are to be tested in this data. We hope to submit an abstract on lung function trajectories for the 2017 European Respiratory Society Congress (abstract deadline anticipated February 2017) and to prepare the relevant manuscript in spring 2017.

4. **Rationale:**

Chronic obstructive pulmonary disease (COPD) is the third leading cause-of-death worldwide (1). Despite large reductions in the prevalence and intensity of smoking over the last 50 years, age-adjusted COPD mortality has remained high, and most cases of incident COPD occur now in former smokers and persons who have never smoked (2-7). Beyond smoking, occupational and environmental exposures, gene variants, and early life factors have been identified as risk factors for COPD, yet considerable variability in COPD incidence and prognosis remains unexplained (8). Prediction and prevention of COPD are particularly important given the lack of medical therapies proved to reduce COPD-related mortality (8, 9).

An accelerated rate of decline in lung function, which may be caused by smoking, may lead to the development of COPD (7); however, there is increasing recognition that low
lungs function in early adulthood may be another trajectory leading to COPD (10-12).
This has encouraged reconsideration of classical paradigms in order to identify novel risk
factors as well as strategies for primary prevention of chronic lung diseases – which,
besides smoking cessation and avoidance, are currently lacking (13).

Furthermore, definitions for COPD are now being reexamined. COPD is currently
defined functionally by airflow limitation on spirometry that does not fully reverse with
bronchodilators, together with respiratory symptoms (14). Virtually all clinical trials of
medical therapies to date use this definition. Nonetheless, recent research shows that
COPD-like symptoms and exacerbations may occur with preserved lung function (15,
16), warranting the identification and investigation of alternative COPD phenotypes.

We therefore propose to assess lung function trajectories and their relations to known
clinical risk factors (e.g., pack-years, COPD genetic risk scores) as well as to alternate
case definitions and clinically important outcomes of COPD (e.g., symptoms, respiratory
events). We plan to use data from six epidemiologic cohorts which collected longitudinal
spirometry data. These data have been harmonized and pooled as part of the NHLBI
Pooled Cohorts Study (NIH/NHLBI R21-HL121457, R21-HL129924, K23-HL130627),
thereby providing a large, multiethnic sample of smokers and never-smokers from which
we aim to draw inferences generalizable to the US population.

5. Main Hypothesis/Study Questions:

1. To describe and discriminate between different lung function trajectories in a US
general population-based sample. In particular, we will:
   a. Determine the prevalence of four lung function trajectories identified in
      recent literature based upon cross-tabulation of starting values for FEV1
      percent predicted (>=80% versus <80%) versus incidence of
      spirometrically-defined COPD over follow-up (yes or no).
   b. Compare the sensitivity of these results to different spirometric starting
      values and different case definitions for incident COPD.
   c. Compare these results based upon these a priori groupings of trajectories
      to those obtained from unsupervised clustering approaches.
   d. Characterize population mean changes in lung function over time,
      including within trajectory groupings, using generalized estimating
      equations and generalized linear mixed models.

2. To characterize risk factors for different lung function trajectories, including
   evaluation of:
   a. Empirical cutoffs for pack-year exposure corresponding to marked
      accelerations in lung function decline
   b. Empirical cutoffs for “age-related” accelerations in lung function decline
   c. The contribution of genetic risk scores for COPD and maximally attained
      lung function in predicting lung function trajectory
3. To test for the clinical significance of different lung function trajectories, as defined in (1), with respect to:
   a. Respiratory symptoms
   b. Initiation of inhaler therapies
   c. Severe obstructive lung disease events (SOLE) – defined as hospitalizations/deaths with COPD, chronic bronchitis, emphysema, or asthma as the primary discharge diagnosis code or underlying cause of death (17)
   d. All-cause mortality

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).

Data
We propose to use six 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. Multiethnic Study of Atherosclerosis (MESA)

We will use the following data:
- All spirometry measures, including QC variables, from all available exams
- Symptoms: self-reported respiratory symptoms including dyspnea, wheeze, cough
- Events: occurrence and time-to-event from study baseline for hospitalizations/deaths with international classification of disease (ICD) codes for asthma, COPD, chronic bronchitis, emphysema. SOLE, or hospitalizations and deaths attributable to chronic lower respiratory diseases, will be identified via ICD codes (ICD-9 490-493, 496, 506.4; ICD-10 J40-J45). Events will be subclassified by code position (primary diagnosis code or underlying cause of death versus any code position) and sub-type (e.g., COPD versus bronchitis). Adjudicated exacerbations and deaths due to CLRD (available in HABC, MESA, and HCHS/SOL) will be evaluated as alternate definitions of SOLE (17).
- Socio-demographics: age, sex, race/ethnicity, insurance status, socioeconomic status
- Anthropometric: height, weight, BMI, waist-to-hip ratio, waist circumference
- Smoking: smoking status, cigarettes per day, pack-years, pipe use, cigar use
Medical history: history of COPD, asthma, medications (including inhalers, steroids, HRT, anti-hypertensives, statins, antiplatelets)
Other exposures: occupational history/exposures, air pollution exposure (as available)

Of note, data from all cohorts have already been obtained, harmonized, and pooled under the auspices of approved paper/ancillary study proposals relating to other specific biological hypotheses.

All data will be harmonized prior to pooling as described in a separate paper proposal submitted simultaneously.

Analytic Plan

In order to describe and discriminate between different lung function trajectories, longitudinal spirometry data will be analyzed using several different approaches:

- In an effort to replicate the findings of Lange et al, we will employ the same approach, which was to stratify the sample by starting lung function values and incidence of spirometrically-defined COPD, and then to define annual decline in lung function in each stratum according to simple slopes (i.e., starting – ending FEV1 / years elapsed). As in Lange et al, COPD will be defined as GOLD Grade 2 or higher (10). GOLD grade 2 is defined by a ratio of FEV1 to forced vital capacity (FVC) of less than 0.70 and an FEV1 of less than 80% of the predicted value. We will test the sensitivity of these findings to alternate spirometric thresholds (e.g., different starting values for FEV1; starting values for FVC; starting values for FEV1/FVC) and different case definitions for incident COPD (e.g., alternative spirometric definitions, incident respiratory symptoms, incident SOLE).

- We will compare the *a priori* groupings of trajectories provided in Lange et al to those obtained by unsupervised clustering approaches within a finite mixture model with time-varying covariates (e.g., PROC TRAJ), using incidence of spirometric thresholds, respiratory symptoms, and SOLE as outcomes to inform the clusters (19). Results will be compared with respect to consistency as well as construct and predictive validity vis-à-vis smoking history, symptoms and incidence of SOLE.

- To characterize population mean changes in lung function over time, including within trajectory groupings, we will analyze longitudinal spirometry data using generalized estimating equations and generalized linear mixed models. In the pooled data, we will use a cohort-specific unstructured covariance matrix, akin to a “stratified MANOVA” approach. Empirical best linear unbiased predictions will be estimated. Splines will be fit to characterize non-linearity. To account for
attrition, we will explore inverse probability weighting approaches conditioning-on-being-alive, as previously accomplished in ARIC (18).

To characterize risk factors for different lung function trajectories, we will:
- Examine to what extent participant characteristics, such as pack-year history or COPD genetic risk score, are informative regarding predicted lung function trajectory
- Identify “high-risk” thresholds for pack-years, age, BMI, and other participant characteristics with respect to rapid lung function decline and/or incident COPD

To test for the clinical significance of different lung function trajectories, we will:
- Treat “trajectory” (as defined in Main Study Question # 1) as a risk factor for incident COPD outcomes (e.g., spirometrically-defined COPD, symptoms, SOLE, and all-cause mortality)
- Compare to what extent “historical” lung function measurements are informative with respect to risk of COPD outcomes after taking the most recent lung function measurement into account, using generalized linear models with time-varying covariates.

Statistical analyses will be performed in R or SAS, Version 9.3.

7.a. Will the data be used for non-CVD analysis in this manuscript? ___ ___ Yes ___ No

b. If Yes, is the author aware that the file ICTDER03 must be used to exclude persons with a value RES_OTH = “CVD Research” for non-DNA analysis, and for DNA analysis RES_DNA = “CVD Research” would be used? ___ ___ Yes ___ No
(This file ICTDER has been distributed to ARIC PIs, and contains the responses to consent updates related to stored sample use for research.)

8.a. Will the DNA data be used in this manuscript? ___ ___ Yes ___ 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.cscu.unc.edu/ARIC/search.php
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)?

The authorship group for this proposal has several approved proposals that test non-overlapping, specific biological hypotheses in the harmonized and pooled data (AS 2013.04, 2014.41, 2016.09).

Mirabelli MC, et al, Respiratory Medicine, 2016 provides a method to address the effects of attrition on longitudinal lung function analyses in ARIC. While the lead author is currently working at the CDC, a number of co-authors on this study are included in our proposal, and we plan to explore applying these methods developed in ARIC across all cohorts.

11.a. Is this manuscript proposal associated with any ARIC ancillary studies or use any ancillary study data? ______ Yes     ______ 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.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.