1a. Full Title: Dietary Dairy Intake, Lung Function, and Chronic Obstructive Pulmonary Disease (COPD) in the Atherosclerosis Risk in Communities (ARIC) Study.

b. Abbreviated Title (Length 26 characters): Dairy Intake, Lung Function, and COPD

2. Writing Group: Matthew B. Schabath, Alvaro Alonso, Gerardo Heiss, and Jennifer Nettleton

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

First author: Matthew B. Schabath
Address: Division of Epidemiology
UTHSCH School of Public Health
1200 Herman Pressler Suite #617
Houston, Texas 77030

Phone: 713-500-9216 Fax: 713-500-6264
E-mail Matthew.B.Schabath@uth.tmc.edu

Corresponding/senior author (if different from first author correspondence will be sent to both the first author & the corresponding author):
Address: Matthew B. Schabath
Address: Division of Epidemiology
UTHSCH School of Public Health
1200 Herman Pressler Suite #617
Houston, Texas 77030

Phone: 713-500-9216 Fax: 713-500-6264
E-mail Matthew.B.Schabath@uth.tmc.edu

3. Timeline:

   Manuscript Revision: Dec 2008
   Manuscript Submission: Jan 2009
4. Rationale:

Non-malignant respiratory diseases have considerable impact on public health. In the US, chronic obstructive pulmonary disease (COPD) is currently the fourth leading cause of death, affects ~11% of the population, and is a leading cause of hospitalization among adults. Despite the magnitude of the problem, preventive and therapeutic strategies for COPD are limited. Although smoking is the major environmental risk factor for lung function impairment and COPD, only a fraction of smokers experience accelerated lung function impairment and actually develops COPD (1,2). Diet may play an important role in the preventing or delaying the rate of subclinical lung function decline and the development of COPD. Previous studies, including ARIC collaborators, have shown dietary factors were putative protective factors of lung function and COPD. Kan et al. (3) found that dietary fiber was protective for COPD and reduced lung function after taking into account intakes of antioxidant micronutrients. Furthermore, the authors found that micronutrient intakes were also protective. These findings confirm other findings on chronic bronchitis from a cohort of adults in Singapore (4). In the ARIC cohort, Shahar et al. (5) found that intake of omega-3 fatty acids was related to reduced lung function and decreased risk of COPD. These limited data provide evidence for a role of diet, lung function, and COPD.

The present proposal plans to explore the association of dietary intake of dairy, lung function parameters, and COPD phenotypes. Dairy products are primary sources of vitamin D, magnesium, and calcium and may be associated with lung function and risk of COPD, yet there are limited epidemiological data that have explored this association. In animal models and human lung-specific in vitro studies, vitamin D has been suggested to play an important role in lung development, lung function, regulation of gene expression, and stimulation of surfactant synthesis (6-10). Magnesium may also play important because of its important role in many biological mechanisms including acting as a cofactor in enzyme activation reactions requiring adenosine triphosphate, acting as a bronchodilator of airway smooth muscle, inhibition of cholinergic neuromuscular transmission, and stabilization of mast cells and T lymphocytes (11). Calcium is an important mediator by which oxidative stress may modulate signal transduction pathways and high calcium intake influences the circulation of vitamin D metabolites. We therefore propose to examine the association of dairy intake with lung function parameters and COPD phenotypes. Stratified analyses will be explored to test for interactions and to look for highly susceptible subgroups.

5. Main Hypothesis/Study Questions:

1. Determine the relationship between dietary intake of dairy and COPD phenotypes and lung function parameters. A cross-sectional analysis, including a case-comparison analysis of the COPD phenotypes, will be performed at the baseline visit. Multivariable linear regression models will be used to assess the association diary and the lung function parameters, while multivariate logistic regression will be used to assess the association dairy and the COPD phenotypes. **We hypothesize that high intake of dairy will be associated with higher lung function parameters and lower risk of COPD phenotypes.**

2. Perform stratified analyses from Aim 1 to test for interactions and highly susceptible subgroups (e.g., smoking, ethnicity, and BMI) for the relationship between dietary intake of dairy and COPD phenotypes and lung function parameters.

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:** The proposed analysis will be a cross-sectional analysis utilizing the clinical, epidemiologic, dietary, exposure, respiratory symptom, and lung function data from visit 1 (i.e. baseline).
**Inclusion/exclusion:** The non-CVD restrictions, ethnic group and missing data exclusion criteria will be used. Participants reporting extreme energy intake or with otherwise implausible dietary responses will also be excluded (upper and lower 1% of the distribution, using the “INCLUDE” variable).

**Outcome:** The main outcomes for these analyses will be COPD phenotypes and lung function parameters. We have 30 different LF parameters (Table 1) on all available study subjects. Gender- and race-specific equations, that include height and age, will be utilized to calculate percent predicted, where appropriate.

---

**Table 1. List of lung function parameters**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>FEV1 % Predicted</th>
<th>FVC liters</th>
<th>FEV1 liters</th>
<th>PEFR</th>
<th>FVC Predicted liters</th>
<th>FEF25% Predicted</th>
</tr>
</thead>
<tbody>
<tr>
<td>FEV1 % Predicted</td>
<td>FEV1 liters</td>
<td>PEFR</td>
<td>FVC Predicted liters</td>
<td>FEF25% Predicted</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FVC liters</td>
<td>FEV3 liters</td>
<td>FEF25%</td>
<td>FEV0.5 Predicted liters</td>
<td>FEF50% Predicted</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FEV1/FVC</td>
<td>FEV5 liters</td>
<td>FEF50%</td>
<td>FEV1 Predicted liters</td>
<td>FEF75% Predicted</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FEV1/FEV6</td>
<td>FEV3/FEV6</td>
<td>FEF75%</td>
<td>FVC % Predicted</td>
<td>FEF25%-75% Predicted</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FEV0.5 liters</td>
<td>FEV0.5/FVC</td>
<td>FEF25%-75%</td>
<td>FEV3 Predicted liters</td>
<td>FEF50% Predicted</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FVC % Predicted</td>
<td>FEV/FVC</td>
<td>Time to best FVC secs</td>
<td>PEFR Predicted</td>
<td>FEF25%-75% Predicted</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

The criteria in Table 2 will be used to define the COPD phenotypes. To eliminate the possibility of reversible concomitant disease (i.e., asthma), we will exclude individuals with a self-reported history of asthma. COPD will be defined using PFT data and modified GOLD criteria i.e. FEV1/FVC < 0.70 & FEV1 < 80% predicted (Table 2). We will utilize clinically relevant physician diagnosed self-report data to discern between emphysema and chronic bronchitis. Chronic bronchitis by physician diagnosed self-report (i.e., CBSR) will be defined as individuals who self-report a physician diagnosis of chronic bronchitis or who self-reported a history of chronic cough or phlegm production for 3 or more months for 2 or more years. The CBSR phenotype will be particularly useful since these criteria are quite similar to the clinical criteria used to diagnosis chronic bronchitis. Additionally, not all individuals who have chronic bronchitis will have a FEV1/FVC < 0.70, so this phenotype will not exclude individuals who have chronic bronchitis but a preserved FEV1/FVC. Classification in this manner will also allow us to analyze two subgroups of CBSR, i.e., those with and without a preserved FEV1/FVC. Emphysema (i.e., EMPH) will be defined as individuals who self-report a physician diagnosis of emphysema. We found that 77% of individuals who had a physician diagnosed history of emphysema had a FEV1/FVC of < 0.70 which suggests that physician diagnosed emphysema has a relatively high concordance with the PFT criteria. Overall, by using these phenotype criteria to discriminate between emphysema and chronic bronchitis with and without “normal” PFTs, we will have the potential to explore these disease processes separately and determine if specific diseases segregate for a given genetic association. Normal (i.e., NORM) will be defined as individuals who do not meet any of the phenotype criteria and have a FEV1/FVC ≥ 0.70 and FEV1 ≥ 80% predicted.

---

**Table 2. Criteria to Define the COPD and "Normal" Phenotypes in ARIC**

<table>
<thead>
<tr>
<th>Phenotype</th>
<th>Abbreviation</th>
<th>Criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>COPD defined by LF</em></td>
<td>COPD</td>
<td>FEV1/FVC &lt; 0.70 and FEV1 &lt; 80% predicted</td>
</tr>
<tr>
<td><em>Physician diagnosed emphysema</em></td>
<td>EMPH</td>
<td>Self-report physician diagnosed emphysema</td>
</tr>
<tr>
<td><em>Physician diagnosed chronic bronchitis with “non-normal” LF</em></td>
<td>CBPFT-</td>
<td>CBSR individuals with a FEV1/FVC &lt; 0.70</td>
</tr>
<tr>
<td>Chronic bronchitis by self-report with “normal” PFT</td>
<td>CBPFT+</td>
<td>CBSR individuals with a FEV1/FVC ≥ 0.70</td>
</tr>
<tr>
<td>Chronic bronchitis CB by self-report</td>
<td>CBSR</td>
<td>Self-report physician diagnosed CB or self-reported history of chronic cough or phlegm production for 3 or more months for 2 or more years</td>
</tr>
<tr>
<td>Normal</td>
<td>NORM</td>
<td>Are not classified above and FEV1/FVC ≥ 0.70 and FEV1 ≥ 80% predicted</td>
</tr>
</tbody>
</table>
Main Exposures:
Self-reported dairy intake at baseline (from interview-administered, 66-item food frequency questionnaire [FFQ]). We will characterize dairy food intake as high-fat and low-fat by combining individual line items (variable name given in parentheses) from the baseline FFQ responses as follows:

High-fat dairy (in units of servings/day) = high-fat milk (hfmilk01) + ice cream* (icecream01) + high-fat cheese (hfchs01)

*Because ice cream is unique in its nutrient contribution (i.e., lower in protein, higher in added sugar), we will also characterize high-fat dairy intake excluding ice cream (high-fat milk and cheese only).

Low-fat dairy (in units of servings/day) = low-fat milk (lfmilk01) + yogurt (yogurt01) + low-fat cheese (lfchs01)

Associations with individual dairy components (e.g. Vitamin D) will also be explored to determine which, if any, individual food item drives observed associations. High-fat dairy and low-fat dairy intake will be modeled in quartiles based on the sample intake distribution. \( P \) for difference among percentiles values (means or odds ratios) will be determined from the F-test indicating inequality among the percentile categories (with \( n \) of freedom).

Because diet is notoriously measured with error, we will explore the effects of bias due to this error. If we feel the degree of bias is large, we will explore methods to correct for some degree of the error in reported dietary intake.

Other variables of interest: Environmental exposures (smoking, SES, occupation, physical activity, diet, etc), prior medical history (e.g., prior respiratory disease), and covariate data will be required.

Summary of data analysis: After the completion of data merging, validation, and query resolution, a series of standard exploratory data examination steps will be performed before the formal analytical methods are applied. Where appropriate, parametric and non-parametric analyses will be used to describe and characterize the data. Where required, Pearson’s \( \chi^2 \) test or Fisher’s exact test will be used to test for differences in the distribution of categorical data such as sex, ethnicity, SES factors, and smoking status. Student’s \( t \) test will be used to test for differences, when required, for continuous variables such as age and pack-years smoked. A nonparametric alternative, the Wilcoxon rank-sum test, will be considered when data are markedly non-normal. Continuous variables will be summarized by descriptive statistics, tests for normality will assess the distribution of the data, and graphical analyses will also be explored. If necessary, variables will be transformed. Unless noted otherwise, all statistical analysis will be performed using the Intercooled Stata 10.0 statistical software package (Stata Corp., College Station, TX).

**COPD Phenotypes:** Multivariable logistic regression analyses will largely be used to assess the association between diary intake and COPD phenotypes. ORs and 95% confidence intervals will be calculated as an estimate of the relative risk. Each COPD phenotype will be coded as “1” and the NORM phenotype will be coded as “0”. The dietary data will be analyzed by percentiles (e.g. quartiles, quintiles, etc). Multivariable logistic regression analysis will be conducted to assess if diary intake is independently associated with risk of the COPD phenotypes adjusting for potential confounders, including age, sex, ethnicity, education, smoking status, pack-years of smoking, history of CHD, BMI, other dietary factors, and any other statistically significant or biologically/clinically relevant variables. Additionally, formal analytical methods will be applied to determine if any of these or other factors are a confounder. If any factor is deemed a confounder, we will control for the confounding in the multivariable analysis. Because COPD is a multi-causal disease whereby many
factors may contribute to and modify risk, careful consideration will be followed to assess the impact of existing medical conditions, such as atopy, obesity, cardiac disease, stroke, etc. Using stratified analyses we will determine if these factors are potential effect modifiers. If any factor is considered to be an effect modifier, we will decide, based on sample size considerations whether to exclude individuals with the specific factor or report strata-specific effects.

**Lung Function Parameters:** The association between the diary and lung function parameters will be assessed using a multivariable linear regression adjusting for age, sex, ethnicity, smoking status, pack-years of smoking, history of CHD, BMI, other dietary factors, and any other statistically significant or biologically/clinically relevant variables. Again, diary intake will be analyzed by percentile and the multivariable linear regression will be used to obtain the adjusted means for each level of percentile. The F-test will be used to determine if there is a statistically significant difference across the percentiles.

**Any anticipated methodological limitations or challenges:** At this point we do not anticipate any substantial methodological challenges. However, it should be noted that the cross-sectional design of this study is a limitation since we cannot assess temporal changes of diet, lung function, or incident COPD. Given the measurement error in lung function testing, there is too much noise to detect longitudinal changes with two measurements < 4 years apart. Therefore, we will have limited ability, if any at all, to detect longitudinal changes if we also applied visit 2 to these analyses. However, we can take advantage and maximize the available data and study design by the approach described above.

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 ICTDER02 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 ICTDER02 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

b. If yes, is the author aware that either DNA data distributed by the Coordinating Center must be used, or the file ICTDER02 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.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)? None at this time.

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.cscu.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 for rationale


