Title: Comparing lung cancer risk prediction among men and women: The NHLBI Pooled Cohorts Study.

Authors:
Jenna Gibilaro, Columbia University, New York NY, USA, jmg2313@cumc.columbia.edu
Pallavi Balte, Columbia University, New York, NY, USA, ppb2119@cumc.columbia.edu
Patricia A. Cassano, PhD, Cornell University, Ithaca, NY, USA, pac6@cornell.edu
Benjamin M. Smith, MD, MS, Columbia University, New York NY, USA, bs2723@cumc.columbia.edu
David Couper, PhD, University of North Carolina, Chapel Hill, North Carolina, USA, david_couper@unc.edu
David R. Jacobs Jr., PhD, University of Minnesota, Minneapolis, MN, USA, jacob004@umn.edu
Ravi Kalhan, MD, Northwestern University, Chicago, IL, USA, RKalhan@nm.org
George T. O’Connor, MD, Boston University, Boston, MA, USA, goconnor@bu.edu
Joseph E. Schwartz, PhD, Columbia University, New York, NY, 10032, jes2226@cumc.columbia.edu
Sachin Yende, MD, University of Pittsburgh, Pittsburgh, PA, USA, Yendes@upmc.edu
Paolo Chaves, MD, Florida International University, Miami, FL, USA, pdehendo@fiu.edu
Wendy B. White, PhD, Tougaloo College, Tougaloo, MS, USA, wendywhite2001@yahoo.com
Elizabeth Oelsner, Columbia University, New York, NY, USA, eco7@cumc.columbia.edu

Introduction

More men and women die from lung cancer than any other cancer in the United States (1). Despite improved survival over the last few decades in many types of cancer, lung cancer mortality remains comparatively high with an estimated 5% 5-year survival rate (2).

Smoking has long been recognized as the major, modifiable risk which leads to lung cancer in both sexes (3). Indeed, current clinical prediction models for lung cancer rely heavily on demographic data and smoking status, but include few if any clinical variables (4-6). Increased lung cancer incidence and mortality in men versus women observed in the past has been attributed to differences in smoking patterns, such as younger onset of smoking and more cigarettes per day in men (7, 8). While smoking behaviors have decreased in both men and women, the decrease is more pronounced in men; accelerated smoking cessation in women lagged two decades behind that of men (9). These differences in smoking habits may explain a more accelerated drop in incident lung cancer in men versus women in the past (2). However, smoking habits between men and women are now quite similar (7, 8), and yet there is currently an unexplained increased incidence of lung cancer in women (10, 11), particularly in young women versus young men who are non-Hispanic white and Hispanic (12). This unexpected increase in incident lung cancer among young women versus young men constitutes a major, unanswered public health question.

Beyond smoking, differential risks of developing lung cancer among men and women may be due to a number of factors that are not measured in large administrative and survey-based datasets, nor in large
cancer screening cohorts. Prior work has found differences in risk of respiratory disease and responses to treatment in women over men (13-15). This may be due to different environmental exposures, such as household air pollution, which has been has been associated with COPD and lung cancer risk (16, 17) and may disproportionately impact women in domestic roles (18). From a biological standpoint, both estrogen and progesterone receptors found in human lung tissue have been implicated in the pathogenesis of neoplastic and non-neoplastic lung diseases (19).

Hence, additional research into potential risk determinants and predictive factors for lung cancer in men and women are urgently needed. We therefore propose to use a large, highly-characterized, US general-population based sample to test associations between individual characteristics and lung cancer incidence and mortality, and to assess for potential effect modification by sex. In our analysis we will include well established risk factors for lung cancer, such as age and smoking, and expand to examine less explored potential risk factors in a sample that, like the contemporary US population, includes a large proportion of low-intensity and never-smokers: longitudinal lung function, respiratory symptoms, markers of inflammation, co-morbid medical conditions beyond COPD, age at menopause and use of hormonal replacement therapy (HRT).

Main study questions and hypotheses

1. Current lung cancer prediction models will perform better in men versus women, and there will be evidence for effect modification by sex for age and smoking-related risk predictors
2. Lung cancer prediction will be improved, particularly in women, by inclusion of additional clinical factors:
   a. Longitudinal lung function
   b. Respiratory symptoms
   c. Medical co-morbidities
   d. Inflammatory biomarkers
   e. Age at menopause and HRT status

Data
Sample

We propose to use data from four cohorts that have been harmonized and pooled as part of the NHLBI Pooled Cohorts Study (20):

1. Atherosclerosis Risk in Communities (ARIC) Study
2. Cardiovascular Health Study (CHS)
3. Health Aging and Body Composition (Health ABC) Study
4. Multi-Ethnic Study of Atherosclerosis (MESA)

For the primary analyses, we will have 29,064 participants with 439,043 person-years of follow-up for lung cancer. The composition of the sample will be 54% women, 66% European Americans (EAs), 26% African-Americans (AAs), 44% never-smokers, and 19% current smokers.

Endpoints
- Lung cancer incidence: defined as lung cancer-related hospitalization or mortality, classified by adjudication or administrative criteria (ICD 162-162.9, C33 and C34)
- Lung cancer death: defined as lung cancer-related mortality, classified by adjudication or administrative criteria (ICD 162-162.9, C33 and C34)

**Comparative lung cancer prediction models**

<table>
<thead>
<tr>
<th>Name</th>
<th>Components</th>
<th>Derivation sample</th>
<th>Link</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSKCC</td>
<td>Age, gender, years smoked, cig per day, current smoking status, asbestos exposure</td>
<td>CARET trial – mostly heavy smokers, 25% asbestos exposed, 65% male, age 44-75</td>
<td><a href="https://www.ncbi.nlm.nih.gov/pubmed/12644540">https://www.ncbi.nlm.nih.gov/pubmed/12644540</a></td>
</tr>
<tr>
<td>NSLT</td>
<td>Age, gender, smoking status, presence of lung disease, family history, BMI, education level, race</td>
<td>Age 55-74, 59% male, 90% white, either current or former smoking status</td>
<td><a href="https://www.ncbi.nlm.nih.gov/pubmed/21714641">https://www.ncbi.nlm.nih.gov/pubmed/21714641</a></td>
</tr>
<tr>
<td>PLCO-M2012</td>
<td>Modified to compare to NSLT: Age, education, BMI, family history, COPD, current vs former smoker, pack-years, duration of smoking, quit time, race/ethnicity, h/o other cancer</td>
<td>PLCO, except excluded patients who have never smoked: Ages 55-74, 85% white, all current and former smokers, 49% men</td>
<td><a href="https://www.nejm.org/doi/full/10.1056/nejmoa1211776">https://www.nejm.org/doi/full/10.1056/nejmoa1211776</a></td>
</tr>
</tbody>
</table>

All components of the above models will be included as variables in our planned analysis.

**Additional variables to be tested**

- Socio-demographics: birth year, study cohort
- Anthropometric: height, waist circumference
- Smoking: age at smoking initiation
- Second-hand smoking: duration of exposure (hours/week), place of exposure
- Medical comorbidities: hypertension, diabetes, renal failure, obesity, cancer
- Family history: lung cancer, emphysema, COPD
- Longitudinal lung function: initial lung function (percent predicted FEV1, percent predicted FVC, presence of airflow limitation or restrictive pattern) and change in lung function (declines in FEV1, FVC, and FEV1/FVC)
- Respiratory symptoms: dyspnea and chronic bronchitis, using mMRC criteria
- Serum biomarkers: fibrinogen, lipids
- Hormonal status: age at menopause, age at menarche, HRT status

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.

**Analysis plan and methods**

- Baseline characteristics will be tabulated by sex and lung cancer incidence
- Sex-stratified Cox proportional hazards models will be used to test associations between the abovementioned variables, added sequentially, and lung cancer incidence and mortality, with time-to-event defined as biological age at event, left-truncation for age at study entry, and source cohort treated as a stratum term.
Effect modification will be tested by multiplicative interaction terms in the full sample. The main effect modifier of interest will be sex, but age group, race/ethnicity, smoking history, and birth cohort will also be tested.

In sex-stratified models, the prognostic significance of combinations of variables will be assessed via receiver operating characteristic (ROC) curves and c-statistics. The main comparators will be the MSKCC, NLST, and PLCO2012 models for lung cancer prediction, as summarized above (4-6). Differences in discrimination for more highly adjusted models will be assessed via the ROCCONTRAST procedure. The best prediction model, as identified by ROCCONTRAST, will be compared to the PLCO2012 model with respect to the Net Reclassification Index (21).

Statistical analyses will be performed in R or SAS, Version 9.4.

References


