Ambient Air Pollution and Lung Cancer Risk in the Women's Health Initiative Observational Study

Shilpa N Gowda

A thesis

submitted in partial fulfillment of the

requirements for the degree of

Master of Public Health

University of Washington

2016

Committee:
Joel Kaufman
Parveen Bhatti
Elizabeth A Sheppard

Program Authorized to Offer Degree:

Occupational and Environmental Health Sciences

University of Washington

Abstract

The Association between Air Pollution Exposure and Lung Cancer Risk in the Women's Health Initiative Study Cohort

Shilpa N. Gowda, M.D.

Chair of the Supervisory Committee:
Joel Kaufman, M.D., M.P.H.
Department of Occupational and Environmental Health Sciences

Background: Outdoor air pollution was recently classified as a human carcinogen by the International Agency for Research on Cancer[1]. However, epidemiologic studies supporting this classification have focused on lung cancer mortality rather than incidence, and most previous studies have included relatively few women. Women, especially those belonging to older age groups and therefore carrying higher risk for lung cancer, have been particularly underrepresented. New studies that address these limitations are needed to further explore and quantify the strength of the association between ambient air pollution and lung cancer. Two major constituents of ambient air pollution include fine particulate matter ($PM_{2.5}$) and nitrogen dioxide (NO_2), for which previously validated,

state-of-the-art geospatial models exist to estimate exposures. NO_2 is considered a proxy measure for traffic- related air pollutants[2].

Methods: We sought to estimate the association between lung cancer and air pollution, using 1,332 lung cancer cases from the 86,146 women in the observational arm of the Women's Health Initiative (WHI) study, a large, U.S.-based cohort of post-menopausal women[3]. We used geospatial models to estimate ambient air pollution exposures, based on participants' residential addresses. We also characterized participants' exposures to traffic-related air pollution by the distances of their residential addresses to primary limited-access highways. Using Cox proportional hazards regression models, we calculated hazard ratios (HRs) for the risk of lung cancer in association with these exposure metrics. We adjusted for several potential confounders, including age, race, body mass index, region of residence (including urbanicity), smoking habits, and socioeconomic status. Furthermore, we conducted exploratory analyses restricted to never smokers, adjusting for second-hand smoke exposures, because we felt that smoking may obscure the true effects of ambient air pollution on lung cancer. Separate exploratory analyses stratified lung cancer outcomes by major histological subtypes. Substantiated in part by prior studies, we felt that that certain histologic subtypes of lung cancer may be more causally related to air pollution than others.

Results: Mean follow-up time of participants was 11 years, with a maximum follow-up time of 15 years. In our primary analyses, our effect estimates were consistent with no increased risk of lung cancer associated with ambient air pollution. For example, when

comparing the highest quartile of $PM_{2.5}$ exposure (>14.59 $\mu g/m^3$) to the lowest quartile (≤10.58 $\mu g/m^3$), we observed a HR of 0.91 [95% Confidence Interval (CI): 0.61-1.34]. Similarly, our effect estimates were consistent with no increased risk of lung cancer associated with shorter distances to roadway. Moreover, our effect estimates were consistent with no increased risk of lung cancer risk among never-smokers, even when adjusting for exposure to second-hand smoke. Additionally, our effect estimates suggested no increased risk of major lung cancer histological subtypes associated with ambient air pollution exposures.

Conclusions: Overall, we did not observe an association between ambient air pollution and lung cancer risk. However, despite the large sample size, our results do not exclude effect sizes seen in some previous studies. Additional years of follow-up and inclusion of participants from the WHI clinical trial arm would enhance study power and may make the results more informative.

BACKGROUND AND SIGNIFICANCE:

A. Global and U.S. Burden of Lung cancer

Lung cancer remains a significant global health problem. It is the most common cancer worldwide and is the leading cause of cancer death among men and women[4]. While smoking remains the leading cause of lung cancer, globally, about 300,000 annual lung cancer deaths occur that are not attributable to tobacco use[5]. The American Cancer Society estimates 158,080 deaths from lung cancer in 2016 in the United States[4]. Likewise, the World Health Organization estimates that there were 1.59 million deaths from lung cancer worldwide in 2012[26]. It is critical that risk factors for non-smoking-

related lung cancers be identified to support prevention efforts for this largely burdensome disease worldwide.

B. Outdoor Air Pollution and Lung Cancer Risk

Multiple large-scale studies from around the world have demonstrated significant associations between outdoor air pollution and lung cancer risk[1]. In the U.S., it is estimated that outdoor air pollution accounts for two percent of lung cancers among non-smokers[6]. While prior studies attest to associations between repeated short-term exposure increases to air pollution and adverse health effects, there is still a relative paucity of data available supporting an association between long-term, low-level exposures to ambient air pollution and lung cancer risk[7].

Air pollution is a complex mixture of many agents. $PM_{2.5}$ refers to air pollution particles measuring less than 2.5 μm in diameter. Prior epidemiologic studies have focused on characterizing exposures to fine particulate matter ($PM_{2.5}$) and have suggested relevance to important health effects besides cancer, including cardiovascular disease, insulin resistance, and adverse birth outcomes[8,9,10]. In addition to vehicular traffic, sources of $PM_{2.5}$ include industry and dust in the environment. Many studies have also focused on estimating exposures to nitrogen dioxide (NO_2), since it is an easily measured criteria air pollutant and is also an effective proxy for urban and traffic-related air pollution exposure[2].

$PM_{2.5}$ may absorb polycyclic aromatic hydrocarbons, which are known carcinogens[11], in addition to transition metals, and other potentially carcinogenic compounds[12,13]. Because of its small size, $PM_{2.5}$ may penetrate deeply into the lungs and induce

inflammation, oxidative stress, fibrosis; it may also exert genotoxic effects in deep airways[12,13,14].

Likewise, NO_2 has been associated with airway epithelial injury[15]. During inflammation by macrophages and other phagocytosing cells, reactive nitrogen species (RNS), such as nitrogen oxide, and reactive oxygen species (ROS) can be formed[16]. Then, RNS and ROS may act as secondary messengers in intracellular signaling cascades that involve the activation of immediate early response genes, c-fos and c-jun, and subsequently activate transcription factors, such as AP-1 and Nf-kB. These transcription factors regulate the expression of a variety of downstream target genes[15]. Therefore, in its induction of a pro-oxidant state, nitrogen dioxide may be able to promote the formation of cancer[15]. Notably, ROS may also be involved in cancer promotion through another mechanism, indirect mutagenesis[15]. There appears to be a dose-response relationship between ROS and its biological effects, and doses capable of inducing tumor promotion are lower than those inducing mutagenesis[15].

C. The IARC Classification of Outdoor Air Pollution as a Lung Carcinogen

In 2013, the International Agency for Research on Cancer (IARC) classified outdoor air pollution, specifically $PM_{2.5}$, as a group I carcinogen for its role in lung cancer pathogenesis[1]. The group I classification designated $PM_{2.5}$ as definitely carcinogenic to humans. Landmark studies relevant to this classification include the American Cancer Society (ACS) study, published in 2009, and the European Study of Cohorts for Air Pollution Effects (ESCAPE) study, published in 2013. Both of these studies were included in the 2014 IARC meta-analysis, described below.

Through a meta-analysis of 18 studies, primarily from China, the IARC found a meta- Relative Risk of 1.09 (95% CI: 1.04 -1.14) for lung cancer, associated with each 10 $\mu g/m^3$ increase in $PM_{2.5}$[17]. The IARC also reported a similar magnitude of association between NO_2 and lung cancer, but the data were less consistent than they were for $PM_{2.5}$[17]. Notably, China has very high pollution levels, and effect estimates from studies conducted in China may not be relevant to other parts of the world, such as the U.S., where levels of air pollution are much lower.

The European Study of Cohorts for Air Pollution Effects (ESCAPE) study used data from 17 cohorts, based in nine European countries, gathering a total of 312,944 participants[33]. Authors reported a HR for lung cancer of 1.18 (95 CI: 0.96 - 1.46)[33]. They found a HR of 1.55 (95% CI: 1.05 - 2.29) for lung adenocarcinoma per 5 $\mu g/m^3$ increase in $PM_{2.5}$[33]. The authors reported no association between lung cancer and nitrogen oxides (the sum of nitric oxide and nitrogen dioxide, or NO_x), with a HR of 1.01 (95% CI: 0.95 - 1.07) per 20 $\mu g/m^3$ increase in nitrogen oxides[33].

A prospective mortality study, conducted by the American Cancer Society (ACS) as part of the Cancer Prevention II study, was also included in the 2014 IARC meta-analysis. This study found a 14 percent (HR 1.14, 95% CI: 1.04 - 1.23) increased risk of lung cancer mortality attributable to each 10 $\mu g/m^3$ elevation in $PM_{2.5}$. This study used air pollution, vital status, and cause of death data collected from detailed questionnaires, from approximately 500,000 adults enrolled in 1982[18].

D. Limitations of Previous Studies on Ambient Air Pollution and Lung Cancer

While prior studies have been conducted to look at the association between ambient air pollution and lung cancer, many suffer from notable limitations. Most have examined lung cancer mortality rather than incidence, have included a relatively small cohort, and have been retrospective in design. The limitations of a few select studies are worth discussing further, because of the roles that these studies played in the IARC classification.

To begin, the American Cancer Society study relied on death certificate data for outcomes assessment and therefore had the limitation of examining mortality rather than incidence. Mortality is a satisfactory proxy for incidence if survival does not vary once an individual contracts disease[19]. However, if survival is variable, then mortality is not a good proxy for incidence. Mortality from cancer depends on factors such as the stage at which cancer is detected and the efficacy of treatment. Studies suggest that survival from lung cancer can be quite variable, depending on timeliness of treatment[20], which may be in part modulated by socioeconomic differences, particularly for earlier stages of disease[21]. Therefore, lung cancer mortality is only a crude proxy for lung cancer incidence. As a result, use of mortality rather than incidence in a study seeking to examine outcomes of lung cancer may introduce bias. Additionally, death certificates often misclassify causes of mortality data, due to factors such as mis-certifications of the medical causes of death and changes to cause of death classification systems[22]. Therefore, the reliance on data from death certificates in this study potentially introduces further bias.

Furthermore, the American Cancer Society study likely suffered from exposure misclassification due to less sophisticated exposure metrics. The American Cancer

Society assigned all subjects in a metropolitan area the same exposure and did not take into account important aspects of exposure variation. In fact, a re-analysis of the American Cancer Society study that examined mortality from particulate air pollution revealed three times greater intraregional variation in $PM_{2.5}$ exposure than had previously been reported for the Los Angeles area[23]. However, the re-analysis found little variation in $PM_{2.5}$ in the New York City area, which may be due to more uniform prevalence of $PM_{2.5}$ in the northeastern United States[23]. The large difference in intra-city variation between these two major metropolitan areas in the United States underscores the importance of using sophisticated exposure metrics that are able to accommodate more fine-scale exposure variations.

Likewise, the aforementioned ESCAPE study also suffered from exposure misclassification. It used a spatially dense stationary monitoring system combined with land-use regression models. The monitoring was done over the course of only a year, and the constructed exposure models estimated exposures from baseline addresses. Importantly, participants' addresses could have changed during the course of the study and thereby alter subjects' true levels of exposure, which the study exposure assessment methods would not have captured. Therefore, the ESCAPE study may not have incorporated all relevant aspects of temporal variation in air pollution exposures among participants. Resulting exposure misclassification could have biased results in the cohort study analysis.

Additionally, the Harvard Six Cities Study was a leading prospective cohort study, examining the long-term effects of $PM_{2.5}$[24]. It followed a cohort of 8,111 adults, randomly sampled between 1974 and 1977 from six cities located in the eastern and

Midwestern United States[24]. Important limitations included its lack of assessment of NO_2 exposures and its assessment of lung cancer mortality outcomes rather than outcomes of lung cancer incidence[24]. Also, each metropolitan area was assigned the same level of exposure. These are important limitations for the reasons previously described.

E. The Importance of Studying Outdoor Air Pollution and Lung Cancer in Postmenopausal Women

Women are more likely to have non-smoking-related lung cancer than are men[25]. In 2012, it was estimated that there were 583,000 new cases of lung cancer and 491,000 deaths from lung cancer among women globally[26]. Relatively few women have been included in previous studies, meaning that these studies have had relatively little power to detect subtle differential effects among women. As such, additional studies of air pollution and lung cancer among women are needed.

Our study's particular focus on postmenopausal women is warranted given that, as with most other cancers, lung cancer occurs most commonly among older individuals. To illustrate this trend, in the United States, about two out of three lung cancers occur among individuals aged 65 years and older, with less than two percent of lung cancers occurring among individuals younger than 45 years old[4]. Data from the National Cancer Institute's Surveillance, Epidemiology, and End Results (SEER) confirm this trend and reveal that the rates of lung cancer increase among menopausal and post-menopausal women, compared to women of younger ages (see image below).

With population aging a continuing problem in many countries, diseases such as cancer that disproportionately affect the elderly are increasingly pressing global public health problems.

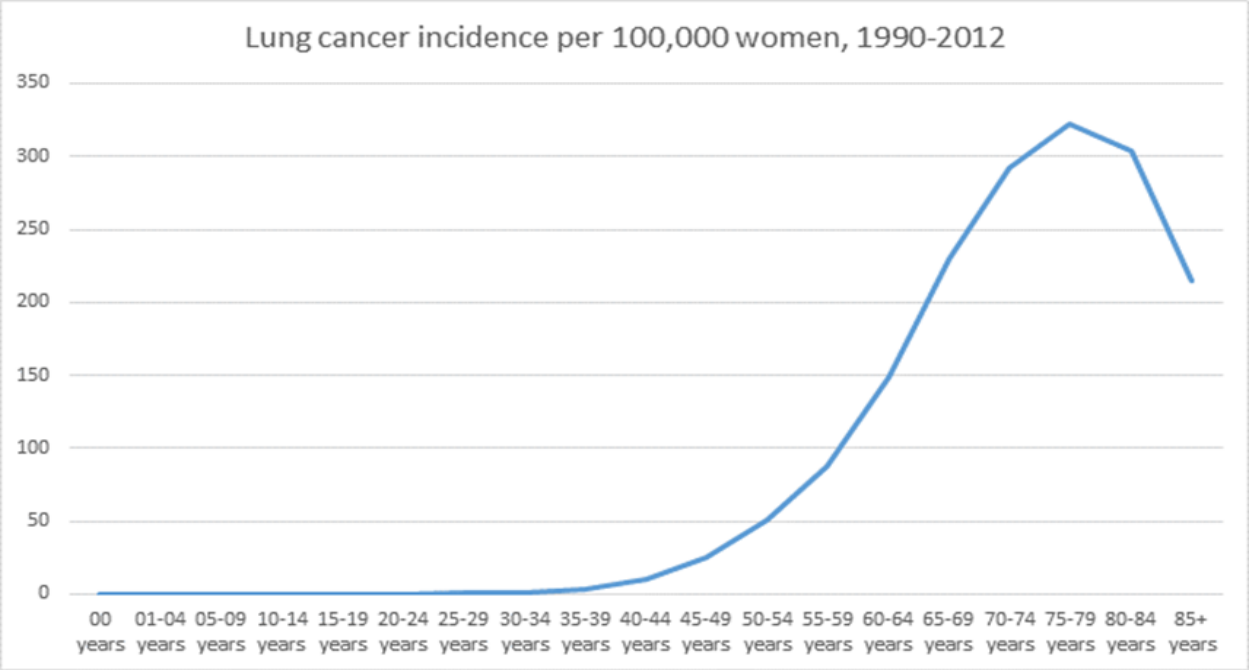

Image 1: Graph depicting incidence of lung cancers per 100,000 women, from year 1990 to 2012 among women of different age groups. Graph constructed from data taken from the Surveillance, Epidemiology, and End Results database.

F. Previous Studies of Air Pollution and Lung Cancer Among Women

The ACS study included a large number of women (126,947) and older individuals. Other notable studies including large numbers of female subjects include the Nurses' Health Study, the Adventist Health Study on Smog (AHSMOG), and the California Teachers' Study. The Nurses' Health Study reported a HR of lung cancer of 1.06 (95% CI: 0.91, 1.25) associated with a 10-µg/m3 increase in 72-month average $PM_{2.5}$, with the

association increasing when analysis was restricted to non-smokers (HR = 1.37; 95% CI: 1.06, 1.77), and when analysis was restricted to lung adenocarcinoma (HR = 1.33; 95% CI: 0.92, 1.93)[27]. On the other hand, the AHSMOG reported inter-quartile range (IQR) increases for PM_{10} exceedance frequencies of 50 $\mu g/m^3$ (RR = 1.21; CI, 0.55-2.66) and 60 $\mu g/m^3$ (RR 1.25; CI, 0.57-2.71). Additionally, the California Teacher's Study reported HRs of 1.45 (95% CI 1.36 – 1.55) and 1.43 (95% CI 1.21 – 1.69) of all-cause and pulmonary mortality, respectively, associated with exposure to $PM_{2.5}$ with an IQR of 6.1 $\mu g/m^3$ with a 30-km buffer. Despite including adequate numbers of women, each of these studies on air pollution and lung cancer risk had important limitations.

The Nurses' Health Study used data from an all-female cohort of 121,700 nurses, who were enrolled in 1976, when they were between 30 and 55 years of age[27]. The study used exposure metrics of stationary monitors to detect levels of air pollution[27]. In urban environments, vehicular traffic is often the major source of air pollution, with peak exposures occurring near, and downwind of, major roadways. As such, distance to roadway has been used as metric of exposure to traffic-related air pollution in many epidemiologic studies and is a form of air pollution exposure assessment called proximity modeling. While proximity models require fewer data and financial resources than other exposure assessment methods, they may fail to account for terrain and meteorological conditions that can modify the exposure[37]. Therefore, substantial exposure misclassification can result.

Moreover, the AHSMOG included 6,338 subjects, of which 4,060 were women[28]. It included both pre- and post-menopausal women, ages 25-79[11]. It reported on other air pollutants but did not examine $PM_{2.5}$ or NO_2.

Similarly, the California Teachers' Study examined 133, 479 pre- and post-menopausal (20-80 years at baseline) female public school professionals[29]. This study examined outcomes of mortality from all causes and from pulmonary disease, rather than mortality from lung cancer specifically[29]. This less specific classification of mortality may have introduced some non-differential misclassification.

DESIGN AND METHODS:

A. Subjects

From 1993 to 1998, 93,676 postmenopausal women (aged 50-79) were recruited to participate in the observational study (OS) cohort of the WHI (http://www.whiscience.org/), a national study with a primary focus to evaluate strategies for preventing cardiovascular diseases, cancer, and osteoporotic fractures. Recruitment was conducted through 40 clinical centers in 24 states and the District of Columbia. Women were screened by telephone to determine eligibility, followed by visits to obtain physical measurements and questionnaire data on demographic, medical, reproductive and lifestyle factors. Participants were asked to complete a questionnaire for updated medical information at least annually to assess health outcomes. Participants in the OS cohort were followed for 8 to 12 years. In 2005, OS participants were invited to enroll in the WHI Extension Study to track health, without intervention, for an additional 5 years.

All cancers were initially identified as part of a medical update questionnaire that was administered at least annually to participants. Those participants reporting a cancer were contacted by mail or phone to get more detailed information about the cancer diagnosis, and then copies of pathology/cytology reports, operative reports, and hospital

discharge summaries were provided to physician adjudicators, who were blinded to exposure status, to confirm or deny the diagnosis, using standardized criteria. Physician adjudicators were required to complete a training process that included reviewing the study protocol and procedures. They were also required to participate in training conference calls held annually and as needed.

B. Covariate Data

Data for factors known or suspected to be associated with lung cancer have been obtained from the WHI baseline questionnaires, interviews and clinical measures. These factors include smoking status; age started smoking; age that smoking was quit; cigarettes per day; years regular smoker; number of years working with a smoker; formerly working with a smoker; currently living with a smoker; living with a smoker after the age of 18; living with the smoker as a child; diagnosis with chronic obstructive pulmonary disease (COPD); age at baseline; ethnicity; years living in state (current at the time of enrollment); U.S. region of residence; highest grade in school; household income; neighborhood socio-economic status index; and residential turnover (AS220; P.I.: Chloe E. Bird).

A z-score variable reflecting neighborhood socio-economic status (NSES) was created for the WHI OS[30]. Factors composing the z-score NSES variable include 1) the percentage of adults aged 25 and older with less than a high school education, 2) percent male unemployment, 3) percent of households with income below the poverty line, 4) percent of households receiving public assistance, 5) percent of households with children headed only by a female, and 6) median household income. To ensure that the index was

comparable from one year to the next, median household income for the years 1991 to 2000 were deflated to 1990 dollars. Further stratification by occupation, such as percent of civilian population 16 years and older with professional, managerial, or executive occupations, was not included in the z-score variable, for this study.

U.S. mean and standard deviation values for each of these six factors were previously generated using the tract-level file. An annual z-score was then generated for each of the six factors, by the following method. The U.S. annual mean for that factor was subtracted from the census tract's annual measure. The difference was then divided by the U.S. annual standard deviation for that factor. The un-normalized index was calculated by taking the annual z-score for median household income, which is positively correlated with NSES, and subtracting from it the annual z-scores for each of the other five variables, which are negatively correlated with NSES. If any of the six selected factors had a missing value, the un-normalized index was set to missing. Finally, the annual un-normalized indices were rescaled such that the values would fall between zero and 100. The rescaled values are the reported annual NSES indices.

C. Air pollution data

We obtained estimates of exposure to ambient air pollution as follows: residential distance to major roadways in meters, $PM_{2.5}$ in micrograms per cubic meter ($\mu g/m^3$), and NO_2 in parts per billion (ppb).

Exposure metrics were estimated using residential address data that were regularly updated during follow-up. Major roads were identified by census code class features A1 through A3[3]. A1 roadways represent primary highways with limited access,

A2 roadways primary roads without limited access, and A3 roadways represent secondary and connecting roads[31]. For $PM_{2.5}$ and NO_2, geocoded subject addresses were linked with exposure estimates derived from geospatial models that incorporated year 2000 EPA air monitoring data and geographic data, including meteorological and traffic volume/pattern data[3]. The spatial models utilized data on 171 distinct Geospatial-Information-System(GIS)-based geographic covariates, chosen from an original set of 265, with multiple indirect measures of traffic, population density, land use, satellite-based vegetative index (NDVI), nearby pollutant emissions derived from emissions inventories, and distances to major sources of pollution. Specific key geographic covariates included 1) population, 2) total emissions of CO, NO_x, PM_{10}, $PM_{2.5}$, and SO_2 in tons per year, 3) percentages of land according to 10 land use categories, 4) summaries of the distribution of the satellite-based MODIS Normalized Difference Vegetation Index (NDVI), 5) measures of impervious surfaces, 6) indirect measures of traffic influences provided by distances to major roads, and 7) distances to commercial zones, airports, smalls shipping ports, railroads, and railway yards[3].

We incorporated spatial smoothing with the models by means of a geostatistical correlation model; and universal kriging allowed the model to make predictions about exposures at arbitrary locations[3]. The spatial models use a cross-validated partial least squares (PLS) approach[3]. Using these models, we assigned each subject in our study the year 2000 estimate of exposure for their residential address for each of the following years under study. These estimates have been used to derive a cumulative average of exposure for the duration of follow-up (i.e., until lung cancer diagnosis or censoring), which is the primary exposure variable of interest.

D. Statistical Analyses

We calculated hazard ratios for the incidence of lung cancer, associated with exposure to $PM_{2.5}$ and NO_2, using Cox proportional hazards regression; length of follow-up was the basic time variable, and we have examined exposures both as continuous variables and quartile categories (based on exposure distributions in the entire population), for our analysis. We stratified the baseline hazard rate by age at enrollment (five-year intervals).

In a secondary analysis, we restricted to never-smoking subjects (n = 47,023; number of cases of lung cancer = 235) and additionally adjusted for two covariates that capture secondhand smoke exposure: lived with smoker as a child and lived with smoker after the age of 18.

We also explored potential histology-specific effects by stratifying on lung cancer subtype. We divided lung cancers into five categories: adenocarcinoma, small-cell lung cancer, non-small cell lung cancer, large cell lung cancer, and squamous cell carcinoma.

RESULTS

After applying exclusions, the final sample size for the study was 86,146 with 1,332 lung cancer cases. The distribution of demographic factors for the WHI OS cohort and incident lung cancer cases is described in Table 1. Subjects were excluded from the study if they had lung cancers at baseline, unknown cancer histories at baseline, missing air pollution data, missing covariate data, or no follow-up time. The mean follow-up

time for the subjects included in our study was 11 years, with a maximum period of follow-up of 15 years.

As Table 1 shows, 92.6 percent of women included in the WHI OS were between 50 and 74 years of age. 85.1 percent of participants were white. 95.1 percent of participants had at least a high school degree, and 42.3 percent of participants belonged to the subgroup having a college degree or higher. 52.5 percent of participants had never smoked. Those who smoked currently and in the past were nearly equally divided between subgroups of smoking histories of fewer than five years, between five and twenty years, and greater than twenty years. 3.7 percent of participants had a history of emphysema, and 7.8 percent of participants had a history of asthma. 39.4 percent of women had body mass indices 18.5 to 24.9, and 34.1 percent were slightly overweight. 66.7 percent of WHI OS cohort participants had a family history of cancer.

Overall, a HR of 0.99 (95% CI: 0.8-1.2) was found for each 10-$\mu g/m^3$ increase in $PM_{2.5}$. A HR of 1.04 (95% CI: 0.93-1.16) was found for each 10 ppb increase in NO_2. Table 2 provides overall and never smoker specific hazard ratios of lung cancer in association with quartile metrics of ambient air pollution exposure. Effect estimates were consistent with no increased risk of lung cancer in association with all ambient air pollution exposure metrics, in the overall study cohort and among non-smokers exclusively.

Table 3 provides the hazard ratios for different lung cancer histological subtypes, in association with quartile metrics of ambient air pollution exposure. Overall, no compelling associations were observed.

DISCUSSION

We observed no associations between lung cancer and ambient air pollution in our study, including when restricting analyses to never-smokers and when stratifying by lung cancer histologic subtype. Recent meta-analyses published in *Environmental Health Perspectives* suggested associations of $PM_{2.5}$ and NO_2 with lung cancer[2,32]. Hamra et al found a meta-relative risk of 1.09 (95% CI: 1.04-1.14) of lung cancer in association with each 10- $\mu g/m^3$ increase in $PM_{2.5}$. In a separate study that he authored, Hamra reported a meta-estimate for the change in lung cancer associated with a 10-$\mu g/m^3$ increase in exposure to NO_2 was 4 percent (95% CI: 1%-8%). Based on 95% confidence intervals from our analyses, our results do not exclude the effects observed by Hamra et al (2014)[39].

To limit the confounding effects of smoking, we conducted additional analyses restricted to never smokers. Previous studies that performed analyses restricting to non-smokers include the Nurses' Health Study. This study included former smokers who had quit at least 10 years prior in the non-smoking analysis. The study reported a hazard ratio of lung cancer of 1.37 (95% CI: 1.06-1.77) in association with $PM_{2.5}$ in this analysis restricting to non-smokers and former smokers who had quit at least 10 years prior, increased from a hazard ratio of lung cancer of 1.06 (95% CI 0.91-1.25) seen in association with $PM_{2.5}$ in the non-restricted analysis.

Furthermore, previous studies have reported varying strengths of associations between air pollutants and different lung cancer subtypes[14], with the strongest

associations observed for lung adenocarcinoma. For example, in the Nurses' Health Study, the strongest association with adenocarcinoma was observed for $PM_{2.5}$, with a hazard ratio of 1.33 (95% CI: 0.92-1.93) in the full cohort and of 1.66 (95% CI 0.81-3.42) in the cohort restricted to non-smokers and former smokers who had quit at least 10 years ago[27]. Similarly, the ESCAPE study reported an increased risk of lung adenocarcinoma, only in association with $PM_{2.5}$ (HR 1.55, 95% CI: 1.05-2.29)[33]. The U.S. has seen an overall rise in adenocarcinomas and decrease in squamous cell carcinomas[33] and there is some debate as to the reason for this trend. It may result from changes in tobacco blends or increasing levels of air pollution[33], though we found no evidence for the latter effect from our study.

As a major strength of our study, the WHI OS cohort allowed us to undertake one of the largest evaluations of air pollution and lung cancer incidence conducted among women to date. It provided a large-scale study population with detailed cancer incidence and covariate data with which to conduct analyses of associations between air pollution and lung cancer. It supported detailed epidemiologic analyses that could be adjusted for potential confounding effects by a large number of covariates (e.g. second-hand smoke exposure). Additionally, the WHI OS cohort dataset enabled examination of outcomes of lung cancer incidence rather than mortality. This provided an advantage over many previous studies, such as the American Cancer Society study, the Harvard Six Cities study, and the California Teachers' Study, which examined outcomes of mortality, which, as previously described, is a crude proxy for lung cancer incidence[18],[24],[29]. Indeed, most previous studies had evaluated cancer mortality, using death certificates. By way of contrast, the classification of outcomes in our study did not rely on death certificate data.

The WHI data on lung cancer incidence outcomes were also very reliable, due to the rigorous classification process described in the methods section. Moreover, the WHI OS outcomes classification was sufficiently detailed to allow for stratification by lung cancer histological subtypes. This stratification allowed us to explore histology-specific associations.

As a major advancement over previous studies, our study used highly validated, state-of-the-art, geospatial models of air pollution to provide detailed estimates of residential ambient air pollution exposure. These were land-use regression (LUR) models that incorporated key information on traffic volumes, land-use, meteorology, and monitoring requirements. In fact, the models used to generate exposure data for our study took advantage of hundreds of geographic covariates, calculated for each location, with monitoring data, to predict detailed exposure levels at any location. The use of hundreds of geographic covariates was made possible by the partial least squares (PLS) regression method. PLS regression finds a small number of linear combinations of the GIS covariates that most efficiently account for variability in the measured concentrations[34]. These linear combinations reduce the covariate space to a much smaller dimension and can then be used as the mean structure in a LUR or universal kriging model in place of individual GIS covariates[34]. For our study, the PLS regression facilitated prediction of two pollutants, $PM_{2.5}$ and NO_2.

Additionally, geospatial models derived from geographic information systems have the advantage over proximity models of being able to account for small-area variations[35]. Notable prior studies were not able to capture such fine-scale exposure variations. For example, the Nurses' Health Study, used nearest monitor exposure

metrics, which did not provide as detailed information on small-area exposure variations as the GIS-derived geospatial models do. Similarly, the American Cancer Society study and the Harvard Six Cities study assigned all subjects in a metropolitan area the same levels of exposure.

On the other hand, our LUR models incorporated geographic data and short-term monitoring information to capture small-area variations in pollution at a fine urban scale. Such capture of small-area variations is important to reduce exposure misclassification. As previously detailed, small-area variations can be significant, and in some areas, intra-city spatial contrast may be as large as the inter-city contrast[36,37]. In order to detect small hazard ratios associated with health effects from ambient air pollution, minimizing exposure misclassification to the greatest extent possible would be important. It is our hope that capturing this extra source of variation has reduced exposure measurement error in our study.

Validation studies support the superiority of LUR models such as ours over other modeling approaches. Such validation studies attempted to characterize the varying abilities of the models to explain the percentage variability (R^2). Collins (1998) compared several Geospatial-Information- System (GIS)-based approaches for modeling of traffic-related air pollution in Huddersfield, UK[36]. These included the moving window approach, ordinary kriging, and dispersion modeling. The moving window approach smooths the spread of highly resolved data on pollution sources (e.g., emissions) to generate a map of concentrations[36]. Kriging spatially smooths residuals developed from exposure prediction models; ordinary kriging as analyzed by Collins differs from universal kriging used in LUR models, because ordinary kriging may not account for the

spatial variation of pollutants to the extent that LUR models do. On the other hand, dispersion modeling is of particular interest when there is a wish to study pollutants admitted from a particular source (e.g., traffic-related air pollution)[36]. Collins found that the best model was derived using LUR ($R^2 = 0.82$), compared to the other three methods, including a moving window approach ($R^2 = 0.67$), ordinary kriging ($R^2 = 0.44$), and dispersion modeling ($R^2 = 0.63$)[36].

On the other hand, our study had some limitations. Residual confounding may have biased our results. A potential confounder that we did not adjust for in our analyses is radon exposure, which is a known, prevalent lung carcinogen. The United States Environmental Protection Agency estimates that one in fifteen homes in the United States has a radon level above the recommended action level of four picocuries per liter of air[38]. EPA data on radon exposures could be linked to the WHI data to enable such adjustments.

Moreover, data for pre-menopausal exposures to air pollution were not available for our study. While stability of residence for at least three years prior to start of the study was a criterion for recruitment into the cohort, WHI OS participants may have resided in different locations prior to their enrollment in the study. Their other places of residence may confer different air pollution exposures than those that the WHI OS dataset alone captures. Therefore, our study could have some exposure misclassification due to this limitation.

Another limitation of our study was the use of year 2000 exposure estimates for each year of follow-up, rather than the use of estimates derived from that year of follow-up. Over the years of follow-up in this study, we would expect spatial, rather than year-

to-year temporal, differences to contribute most to cumulative levels of air pollution exposure. We therefore hope that the lack of temporal variability incorporation in our exposure estimates had minimal impact on our results. Importantly, changes in subjects' residential addresses over the course of follow-up were used to obtain new estimates, based on year 2000 data, for the following years under study. Therefore, our study incorporated some temporal variability in exposure assessment. This incorporation of temporal variability posed an advantage over the ESCAPE study, which, by way of contrast, did not calculate new exposure estimates for subjects who moved during the course of follow-up of the study.

Furthermore, as mentioned previously, the confidence intervals of effect estimates from our study include those reported by Hamra et al. Despite the large size of our study cohort, given the small effect sizes estimated by Hamra et al, a much larger study population than the WHI OS cohort may have been needed to better detect associations that are of similar magnitude to those reported in the Hamra et al papers[39,2]. Additional follow-up of the cohort and addition of participants from the clinical trials arm of the WHI would bolster study power.

Of note, we had not performed sensitivity analyses on our study, as exposure and outcome assessment methods were solid and well validated. Although not clear to us at this time what elements such sensitivity analyses would focus upon, in the future, we may choose to look more carefully for any obvious elements from our study to examine in sensitivity analyses.

In summary, the WHI OS has allowed us to undertake one of the largest evaluations of air pollution and lung cancer incidence conducted among women to date.

Our study had excellent exposure and outcomes assessments. Overall, we did not observe an association between ambient air pollution and lung cancer risk in the WHI OS cohort. However, despite the large sample size, our results do not exclude effect sizes seen in some previous studies. Additional years of follow-up and inclusion of participants from the WHI clinical trial arm would enhance study power. Furthermore, adjustment for radon exposure may decrease residual confounding effects. Future studies that are able to incorporate these changes may make results more informative.

Table 1: Demographic Factors by Lung Cancer Case Status in the WHI OS Cohort

Characteristics	Non-cases	%	Lung Cancer Cases	%
Age at enrollment (years)				
50 – 54	10,991	13	72	5
55 – 59	15,525	19	157	12
60 – 64	18,039	22	277	21
65-69	18,329	22	398	30
70-74	13,620	17	287	22
75-79	6,105	7	141	11
Race/Ethnicity				
White	70,297	85	1,187	89
Black	6,738	8	93	7
Hispanic	3,063	4	15	1
American Indian	358	< 1	5	< 1
Asian/Pacific Islander	1,116	1	15	1
Other	1,037	1	17	1
Education				
Less than high school	1,204	2	12	1
Some high school	2,791	3	59	5
High school diploma	13,334	16	203	15
School after high school	29,949	37	547	41
College degree or higher	34,678	42	504	38
U.S. Region				

Northeast	19,471	24	352	26
South	21,781	26	332	25
Midwest	18,795	23	259	19
West	22,562	27	389	29
Pack Years of Smoking (Categorical)				
Never smoker	43,363	53	235	18
<5	12,391	15	80	6
5 to < 20	11,940	15	192	14
≥ 20	14,915	18	825	62
Age First Smoked by Smoking Status				
Current smoker/age <15	389	1	41	3
Current smoker/age 15–19	2,239	3	170	13
Current smoker/age 20-24	1,506	2	107	8
Current smoker/ age ≥ 25	821	1	33	3
Never smoked	43,363	53	235	18
Past smoker/age < 15	2143	3	65	5
Past smoker/age 15-19	17,371	21	418	31
Past smoker/age 20-24	10,769	13	201	15
Past smoker/age ≥ 25	4,008	5	62	5
History of emphysema				
No	79,529	96	1,196	90
Yes	3,080	4	136	10
History of asthma				
No	76,135	92	1,188	89
Yes	6,474	8	144	11
Body mass index (kg/m^2)				
< 18.5	943	1	17	1
18.5 - 24.9	32,534	39	595	44
25.0 – 29.9	28,166	34	438	33
30.0 – 34.9	13,100	16	177	13
35.0 – 39.9	4,926	6	68	5
≥ 40	2,940	4	37	3
Family History of Cancer				
No	26,381	33	374	30
Yes	52,811	67	892	71

Table 2 –Hazard Ratios of Lung Cancer in Association with Quartile Metrics of Ambient Air Pollution Exposure in the WHI OS

Air pollution exposure	Overall		Non-Smokers	
PM2.5 (μg/m^3)	**HR**	**95% CI**	**HR**	**95% CI**
≤10.58	Ref		Ref	
> 10.58 – 12.39	0.94	0.80, 1.10	0.69	0.47, 1.01
> 12.39 – 14.59	0.94	0.80, 1.10	0.89	0.61, 1.29
> 14.59	0.94	0.79, 1.11	0.91	0.61, 1.34
Per 10 μg/m3 increase	0.99	0.8, 1.2	0.37	0.11, 1.21
NO_2 (ppb)				
≤10.68	Ref		Ref	
> 10.68 – 14.39	1.03	0.86, 1.24	0.99	0.63, 1.53
> 14.39 – 18.19	1.17	0.98, 1.41	1.51	0.98, 2.32
> 18.19	1.06	0.87, 1.29	1.06	0.65, 1.72
Per 10 ppb increase	1.04	0.93, 1.16	1.03	0.57, 1.86
Distance to A1 /A2/A3 Roadway (m)				
>200	Ref		Ref	
50 - <200	1.01	0.90, 1.14	1.03	0.54, 1.99
<50	1.00	0.86, 1.17	1.46	0.69, 3.05

Table 3 - Hazard Ratios of Lung Cancer Subtypes in Association with Quartile

Air pollution exposure	Adenocarcinoma		Small Cell		Non-small Cell	
PM2.5 (μg/m^3)	**HR**	**95% CI**	**HR**	**95% CI**	**HR**	**95% CI**
≤10.58	Ref		Ref		Ref	
> 10.58 – 12.39	0.97	0.76, 1.24	0.81	0.47, 1.37	0.90	0.75, 1.09
> 12.39 – 14.59	1.02	0.80, 1.31	0.56	0.31, 1.02	0.96	0.80, 1.15
> 14.59	0.88	0.68, 1.27	0.71	0.40, 1.27	0.87	0.72, 1.06
NO_2 (ppb)						
≤10.68	Ref		Ref		Ref	
> 10.68 – 14.39	0.94	0.71, 1.26	1.41	0.75, 2.66	0.97	0.78, 1.19
> 14.39 – 18.19	1.11	0.83, 1.48	1.25	0.64, 2.44	1.14	0.93, 1.41
> 18.19	1.05	0.77, 1.42	0.67	0.32, 1.43	0.97	0.77, 1.21
Distance to A1/A2/A3 Roadway (m)						
>200	Ref		Ref		Ref	
50 - <200	1.09	0.90, 1.31	1.38	0.89, 2.14	1.04	0.91, 1.20
<50	0.96	0.75, 1.24	1.39	0.81, 2.37	0.98	0.82, 1.18

Metrics of Ambient Air Pollution Exposure in the WHI OS

Air pollution exposure	Squamous Cell Lung Cancer		Large Cell Lung Cancer	
PM2.5 (μg/m^3)	**HR**	**95% CI**	**HR**	**95% CI**
≤10.58	Ref		Ref	
> 10.58 – 12.39	1.43	0.86, 2.39	0.31	0.15, 0.66
> 12.39 – 14.59	1.27	0.76, 2.14	0.83	0.47, 1.46
> 14.59	1.37	0.81, 2.32	0.65	0.35, 1.23
NO_2 (ppb)				
≤10.68	Ref		Ref	
> 10.68 – 14.39	1.03	0.60, 1.79	0.98	0.45, 2.11
> 14.39 – 18.19	1.12	0.64, 1.95	1.76	0.84, 3.68
> 18.19	1.17	0.65, 2.11	1.42	0.63, 3.17
Distance to A1/A2/A3 Roadway (m)				
>200	Ref		Ref	
50 - <200	0.76	0.52, 1.11	1.08	0.66, 1.76
<50	0.87	0.54, 1.41	0.88	0.45, 1.72

REFERENCES

[1] Loomis, Dana et al. The carcinogenicity of outdoor air pollution. The Lancet Oncology, Volume 14, Issue 13, 1262 – 1263.

[2] Hamra GB, Laden F, Cohen AJ, Raaschou-Nielsen O, Brauer M, Loomis D. Lung Cancer and Exposure to Nitrogen Dioxide and Traffic: A Systematic Review and Meta-Analysis. *Environmental Health Perspectives*. 2015;123(11):1107-1112. doi:10.1289/ehp.1408882.

[3] Sampson PD, Richards M, Szpiro AA, Bergen S, Sheppard L, Larson TV, Kaufman JD. "A regionalized national universal kriging model using Partial Least Squares regression for estimating annual PM2.5 concentrations in epidemiology."
Atmos Environ. 2013 Aug 1;75:383-392. PubMed PMID: 24015108; PubMed Central PMC ID: PMC3763950.

[4] American Cancer Society: Key Statistics for Lung Cancer. http://www.cancer.org/cancer/lungcancer-non-smallcell/detailedguide/non-small-cell-lung-cancer-key-statistics

[5] Alberg AJ, Brock MV, Ford JG, Samet JM, Spivack SD. Epidemiology of Lung Cancer: Diagnosis and Management of Lung Cancer, 3rd ed: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines. *Chest*. 2013;143(5 Suppl):e1S-e29S. doi:10.1378/chest.12-2345.

[6] Causes of Lung Cancer in Nonsmokers. NEWS: Journal of the National Cancer Institute. 2006 May 17; 98(10).

[7] Anderson ML. As the Wind Blows: The Effects of Long-Term Exposure to Air Pollution on Mortality. NBER Working Paper No. 21578. Issued September 2015. Available online: http://www.nber.org/papers/w21578.

[8] Brook et al. AHA Scientific Statement: Particulate Matter Air Pollution and Cardiovascular Disease: An Update to the Scientific Statement From the American Heart Association. Circulation. 2010;121:2331-2378, published online before print May 10 2010, doi:10.1161/CIR.0b013e3181dbece1.

[9] Brook RD, Sun Z, Brook JR, Zhao X, Ruan Y, Yan J, Mukherjee B, Rao X, Duan F, Sun L, Liang R, Lian H, Zhang S, Fang Q, Gu D, Sun Q, Fan Z, Rajagopalan S. Extreme Air Pollution Conditions Adversely Affect Blood Pressure and Insulin Resistance: The Air Pollution and Cardiometabolic Disease Study. Hypertension. 2016 Jan;67(1):77-85. doi: 10.1161/HYPERTENSIONAHA.115.06237. Epub 2015 Nov 16.

[10] Lavigne E, Yasseen AS 3rd, Stieb DM, Hystad P, van Donkelaar A, Martin RV, Brook JR, Crouse DL, Burnett RT, Chen H, Weichenthal S, Johnson M, Villeneuve PJ, Walker M. Ambient air pollution and adverse birth outcomes: Differences by maternal comorbidities. Environ Res. 2016 Jul;148:457-66. doi: 10.1016/j.envres.2016.04.026. Epub 2016 Apr 30.

[11] Flesher JW[1], Lehner AF[2]. Structure, function and carcinogenicity of metabolites of methylated and non-methylated polycyclic aromatic hydrocarbons: a comprehensive review. Toxicol Mech Methods. 2016 Feb 19:1-29.

[12] Borm PJ, Schins RP, Albrecht C. Inhaled particles and lung cancer, part B: paradigms and risk assessment. Int J Cancer 2004; 110: 3–14.

[13] Moller P, Jacobsen NR, Folkmann JK, et al. Role of oxidative damage in toxicity of particulates. Free Radic Res 2010; 44: 1–46.

[14] Dysart MM, Galvis BR, Russell AG, Barker TH. Environmental particulate (PM2.5)

augments stiffness-induced alveolar epithelial cell mechanoactivation of transforming growth factor beta. PLoS One. 2014 Sep 16;9(9):e106821. doi: 10.1371/journal.pone.0106821. eCollection 2014. PubMed PMID: 25226160; PubMed Central PMCID: PMC4167324.

[15] Persinger RL, Poynter ME, Ckless K, Janssen-Heininger YM. Molecular mechanisms of nitrogen dioxide induced epithelial injury in the lung. Mol Cell Biochem. 2002 May-Jun;234-235(1-2):71-80. Review. PubMed PMID: 12162462.

[16] Irigaray, P and Dominique Belpomme. Basic properties and molecular mechanisms of exogenous chemical carcinogens. Carcinogenesis vol.31 no.2 pp.135–148, 2010 doi:10.1093/carcin/bgp252

[17] Loomis et al. The International Agency for Research on Cancer (IARC) evaluation of the carcinogenicity of outdoor air pollution: focus on China. Chin J Cancer. 2014; 33(4):189-196.

[18] Pope CA 3rd, Burnett RT, Thun MJ, Calle EE, Krewski D, Ito K, Thurston GD. Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution. JAMA. 2002 Mar 6;287(9):1132-41. PubMed PMID: 11879110; PubMed Central PMCID: PMC4037163.

[19] Epidemiology for the uninitiated: Quantifying disease in populations. TheBMJ. Available online: http://www.bmj.com/about-bmj/resources-readers/publications/epidemiology-uninitiated/2-quantifying-disease-populations

[20] Nur U, Quaresma M, De Stavola B, Peake M, Rachet B. Inequalities in non-small cell lung cancer treatment and mortality. *Journal of Epidemiology and Community Health*. 2015;69(10):985-992. doi:10.1136/jech-2014-205309.

[21] Berglund A, Lambe M, Lüchtenborg M, et al. Social differences in lung cancer management and survival in South East England: a cohort study. BMJ Open 2012;2:e001048. doi:10.1136/bmjopen-2012-001048.

[22] Foreman KJ, Naghavi M, Ezzati M. Improving the usefulness of US mortality data: new methods for reclassification of underlying cause of death. Population Health Metrics. 2016;14:14. doi:10.1186/s12963-016-0082-4.

[23] Krewski D, Jerrett M, Burnett RT, et al. Extended follow-up and spatialanalysis of the American Cancer Society study linking particulate air pollution and mortality. Health Effects Institute, 2009, 140: 1-154.

[24] Lepeule J, Laden F, Docker D, Schwartz J. Chronic Exposure to Fine Particles and Mortality: An Extended Follow-up of the Harvard Six Cities Study from 1974 to 2009.

[25] Field RW and Withers BL. Occupational and Environmental Causes of Lung Cancer. Clin Chest Med. 2012 December; 33(4):.doi:10.1016/j.ccm.2012.07.001.

[26] Globocan 2012: Estimated cancer incidence, mortality, and prevalence worldwide in 2012. International Agency for Research on Cancer. World Health Organization.

[27] Puett RC, Hart JE, Yanosky JD, Spiegelman D, Wang M, Fisher JA, Hong B, Laden F. 2014. Particulate matter air pollution exposure, distance to road, and incident lung cancer in the Nurses' Health Study Cohort. Environ Health Perspect 122:926–932; http://dx.doi.org/10.1289/ehp.1307490.

[28] Beeson WL, Abbey DE, Knutsen SF. Long-term concentrations of ambient air pollutants and incident lung cancer in California adults: results from the AHSMOG: Adventist Health Study on Smog. Environ Health Perspect. 1998 Dec; 106(12):813-22. PubMed PMID: 9831542; PubMed Central PMCID: PMC1533247.

[29] Ostro B, Lipsett M, Reynolds P, et al. Long-Term Exposure to Constituents of Fine Particulate Air Pollution and Mortality: Results from the California Teachers Study. *Environmental Health Perspectives*. 2010;118(3):363-369. doi:10.1289/ehp.0901181. doi:10.1164/rccm.201012-2082OC.

[30] Chi GC, Hajat A, Bird CE, Cullen MR, Griffin BA, Miller KA, Shih RA, Stefanick ML, Vedal S, Whitsel EA, Kaufman JD. Individual and Neighborhood Socioeconomic Status and the Association between Air Pollution and Cardiovascular Disease. Environ Health Perspect; http://dx.doi.org/10.1289/EHP199.

[31] Census Feature Class Codes- Mississippi Automated Resource Information System (MARIS). Available online: http://www.maris.state.ms.us/pdf/cfcccodes.pdf.

[32] Hamra GB, Guha N, Cohen A, Laden F, Raaschou-Nielsen O, Samet JM, Vineis P, Forastiere F, Saldiva P, Yorifuji T, Loomis D. Outdoor Particulate Matter Exposure and Lung Cancer: A Systematic Review and Meta-Analysis. *Environmental Health Perspectives*. 2014; 122(9).

[33] Raaschou-Nielsen, et al. Air pollution and lung cancer incidence in 17 European cohorts: prospective analyses from the European Study of Cohorts for Air Pollution Effects (ESCAPE).

[34] Bergen S, Sheppard L, Sampson PD, Kim SY, Richards M, Vedal S, Kaufman JD, Szpiro AA. 2013. A national prediction model for $PM_{2.5}$ component exposures and measurement error–corrected health effect inference. *Environ Health Perspect* 121:1017–1025; http://dx.doi.org/10.1289/ehp.1206010.

[35] Brown JS. Emerging applications of Geographic Information Systems (GIS) in community and local mental health research, Journal of Local and Global Health Perspectives 2013:5 http://dx.doi.org/10.5339/jlghp.2013.5

[36] Hoek G, Beelen R, de Hoogh K, Vienneau D, Gulliver J, Fischer P, Briggs D. A review of land-use regression models to assess spatial variation of outdoor air pollution. Atmospheric Environment 42 (2008) 7561-7578.

[37] Jerrett M, Arain A, Kanaroglou P, Beckerman B, Potoglou D, Sahsuvaroglu T, Morrison J, Giovis C. A review and evaluation of intraurban air pollution exposure models. Journal of Exposure Analysis and Environmental Epidemiology (2005) 15, 185-204.

[38] Radon and Cancer. National Institutes of Health National Cancer Institute. Available Online: http://www.cancer.gov/about-cancer/causes-prevention/risk/substances/radon/radon-fact-sheet.

[39] Hamra GB, Guha N, Cohen A, Laden F, Raaschou-Nielsen O, Samet JM, Vineis P, Forastiere F, Saldiva P, Yorifuji T, Loomis D. Outdoor Particulate Matter Exposure and Lung Cancer: A Systematic Review and Meta-Analysis. Environmental Health Perspectives. 2014; 122(9).