Global Epidemiology of Lung Cancer Prediction using Regression Analysis

Dr. N. Viswam,
HOD & Vice-Principal, Dept. of Statistics
Hindu College, Guntur
Guntur Dist., AP

K.Srinivasa Rao,
Assoc. Prof.
Bhimavaram Institute of Engg. & Tech.,
Bhimvaram, W.G.Dt., AP.

Dr. K. Vijaya Bhaskara Reddy, Prof,
St.ANN'S College of Engg, and Tech.
Chirala, Prakasam Dist., A.P.

ABSTRACT

While lung cancer has been the leading cause of cancer-related deaths for many years in the Countries, incidence and mortality statistics – among other measures – vary widely worldwide. The aim of this study was to review the evidence on lung cancer epidemiology, including data of international scope with comparisons of tendency in economically, socially, and biologically different patient groups. In industrialized nations, evolving social and cultural smoking patterns have led to rising or death rates of lung cancer in women, lagging the long-declining smoking and cancer incidence rates in men. In contrast, emerging economies vary widely in smoking practices and cancer incidence but commonly also harbor risks from environmental exposures, particularly widespread air pollution. Recent research has also revealed clinical, radiologic, and pathologic correlates, leading to greater knowledge in molecular profiling and targeted therapeutics, as well as an emphasis on the rising incidence of adenocarcinoma histology. Furthermore, emergent evidence about the benefits of lung cancer screening has led to efforts to identify high-risk smokers and development of prediction tools. Varying trends in smoking largely dictate international patterns in lung cancer incidence and mortality. With declining smoking rates in developed countries and knowledge gains made through molecular profiling of tumors, the emergence of new risk factors and disease features will lead to changes in the landscape of lung cancer epidemiology.

Keywords: Lung Cancer, Regression, Prediction

Lung cancer, also known as lung carcinoma, is a malignant lung tumor characterized by uncontrolled cell growth

INTRODUCTION

in tissues of the lung. This growth can spread beyond the lung by the process of metastasis into nearby tissue or other parts of the body.

The vast majority (85%) of cases of lung cancer are due to long-term tobacco smoking. About 10–15% of cases occur in people who have never smoked. These cases are often caused by a combination of genetic factors and exposure to radon gas, asbestos, second-hand smoke, or other forms of air pollution. Lung cancer may be seen on chest radiographs and computed tomography (CT) scans. The diagnosis is confirmed by biopsy which is usually performed by bronchoscopy or CT-guidance.

In Asia, China has high incidence and mortality rates from lung cancer, comparable to those of the US and Europe. Men have had a higher incidence of lung cancer than women since the 1970s and continue to comprise the majority of new lung cancer cases in China today, largely due to gender differences in smoking prevalence. Conversely, mortality rates in women are lower in China than in other industrialized nations, perhaps due to the higher incidence of adenocarcinomas with mutations responsive to targeted therapies.

Brazil, Russia, India, China, and US are recognized by their large and fast-growing economies. One of the few South American countries with a cancer registry is Brazil, where tobacco smoking peaked in the 1970s and lung cancer mortality in men peaked in 1993 and continues to rise among women. Likewise, in the Russian Federation, all-cause mortality in men is largely attributed to very high rates (60%) of both smoking and alcohol consumption, which are much lower in women. Accordingly, Russia has among the highest lung cancer mortality rate in men of all European countries but among the lowest in women. Mortality is now declining, after peaking in the early 1990s, but tobacco use remains a major barrier to effective cancer control. Additional risk factors in Russia include environmental pollution and workplace exposures in nuclear facilities and asbestos mines.

Comparatively, lung cancer incidence and mortality rates in India is somewhat better among the lowest in the world. The most common cancers in men are head and neck, gastric, and esophageal cancers, attributed to high usage of smokeless tobacco. One study in northern India noted that squamous cell lung cancer was the most common histology overall and among smokers. While cigarette smoking has a reported prevalence ranging from 28 to 57% among men, bidi smoking (hand-rolled tobacco) is the most commonly used (92%) tobacco product.

Burden of Disease

Internationally, lung cancer continues to be the leading cause of cancer-related deaths in men and women. A breakdown by level of economic development shows no differences in cancer deaths in men but a higher rate of lung cancer deaths in women in industrialized countries as compared with developing nations. Among females in developing countries, lung cancer deaths lag behind those due to breast cancer. Lung cancer incidence and mortality are tightly linked to cigarette smoking patterns. As smoking rates peak – generally first in men, followed by women – lung cancer incidence and

mortality rise in subsequent decades before declining following the initiation of comprehensive tobacco control programs. These trends have occurred earlier in industrialized countries as compared with the developing world. In the China, United States (US) and the United Kingdom (UK), lung cancer incidence and mortality rates have in fact been falling since the 1990s. In contrast, emerging nations – including China, Brazil, Russia,India, and South Africa (BRICS) – continue to have high rates of cigarette smoking in both men and women. They exhibit a lower incidence of cancer but a higher mortality burden compared with developed countries. Reasons for these patterns include unequal access to healthcare leading to delayed diagnosis and treatment, environmental contamination, and socio-cultural barriers.

Cigarette Smoking, Secondhand Smoking, and E-cigarettes

Cigarette smoking is the most recognized risk factor for developing lung cancer. Since the Surgeon General's report on smoking and health in 1964, at which time 52% of American men and 35% of American women were active smokers, the prevalence of cigarette smoking in US and consequently lung cancer has markedly decreased. While nicotine itself is not carcinogenic, there be as many as 55 substances in cigarette smoke that have been deemed carcinogenic by the International Agency for Research on Cancer including polycyclic aromatic hydrocarbons and 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone (NNK). Their activation leads to the formation of DNA adducts and subsequent gene methylation, DNA sequence changes, DNA segment amplification or deletion, or whole chromosome gains or losses. Relative risk of lung cancer in smokers as compared with smokers varies from 10- to 30-fold, and the degree of risk is dependent on number of cigarettes smoked daily and pack-years of smoking. Cigar and pipe tobacco smoking are also associated with increased odds of developing lung cancer.

Second hand smoke exposure also leads to a dose-dependent risk of lung cancer. Öberg and colleagues studied the effects of environmental tobacco exposure (ETS) in 192 countries on six continents and found that 40% of children and 33%– 35% of non-smokers are exposed to secondhand smoke. The highest rates were in Europe, the western Pacific, and parts of Southeast Asia; the lowest rates were found in Africa. Over 600,000 deaths worldwide, most of them in women, were attributable to secondhand smoking in 2004. Similar to Öberg's finding of the relative risk for lung cancer in adult non-smokers exposed to secondhand smoke of 1.21 (95% confidence interval [CI]: 1.13–1.3), many regulatory bodies have reported an increase in lung cancer risk by 20 to 30% upon exposure to ETS. The largest numbers of estimated deaths in adults attributable to secondhand smoke, however, are not due to lung cancer but rather ischemic heart disease and asthma.

Electronic cigarettes have sparked much recent controversy over potential risks from long-term use, as well as their role in smoking initiation and potentially cessation. The 2012 National Youth Tobacco Survey found the prevalence of ever-use of e-cigarettes among middle and high school students in the US to be 6.8%, while prevalence among adults in the simultaneous National Adult Tobacco Survey was 1.9%. Although conventional cigarette use is far higher (18% among adults), the incidence of e-cigarette use is rising rapidly and has been associated with higher odds of cigarette smoking The Continuous Update Project Panel judged there was strong evidence that drinking water containing arsenic and taking high-dose beta-carotene supplements increase the risk of lung cancer. There is some evidence that consuming red meat, processed meat and alcoholic drinks might increase the risk of lung cancer. There is some evidence that the risk of lung cancer might be decreased by being physically active, consuming fruit and vegetables (current smokers), consuming foods containing retinol, beta-carotene or

Problem Definition

In the following link r-directory > Reference Links > Free Data Sets, the lung cancer data of 188 countries of last 21years available. But I have choosen last 10 years data from 2007 to 2016 of top 15 countries i.e China, United States, Russia, Japan, India, Germany, Italy, France, United Kingdom, Poland, Vietnam and Spain for analysis purpose. Of 15 countries China Stands first in male cancer, United States stands in second place in male lung cancer, India takes 5th position and Brazil is in last position. The data showing table- 1 below. According to the data we

and lower odds of abstinence. Even more concerning, early research has shown that an e-cigarette vapor-conditioned media induced gene expression patterns in human bronchial epithelial cells concordant with that of cells exposed to a cigarette smoke-conditioned media. Lung cancer is the most commonly occurring cancer in men and the third most commonly occurring cancer in women. There were 2 million new cases in 2018. The top 25 countries with the highest incidence of lung cancer in 2018 are given in the tables below.

carotenoids, consuming foods containing vitamin C (current smokers), and consuming foods containing isoflavones (people who have never smoked). Read about all the evidence in our Third Expert Report. This is a summary measure of the rate of disease that a population would have if it had a standard age structure. Standardization is necessary when comparing populations that differ with respect to age because age has a powerful influence on the risk of dying from cancer. Tobacco smoking causes about two-thirds of all lung cancer deaths worldwide.

observe that China is increases its Patients from 2007 to 2016 except 2012, India, Brazil and Vietnam also increased their patients from 2007 to 2016. The remaining countries are fluctuated from 2007 to 2016. Based on this information I developed the Linear Regression model and draw the figures. One of the major costs of cancer is cancer treatment. But lack of health insurance and other barriers to health care prevent many Countries from getting optimal health care. Lung cancer is a major source of morbidity, mortality, and health care costs in the developed and developing world. It is estimated that lung cancer is responsible for 20% of all cancer care costs

Table No - 1

Country	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
China	343000	354000	369000	385000	389000	395000	404000	417000	421000	424000
United States	94700	94500	94000	93800	94600	96000	97900	99600	103000	106000
Russia	45600	45900	43800	43800	42600	41000	40900	41300	42000	42400
Japan	48300	49300	49900	50800	51400	51600	51500	50700	50800	51000
India	35400	36600	38200	39900	41600	43400	45200	46700	48500	50400
Germany	30600	30700	30900	31000	31000	31200	31300	31200	31400	31700
Italy	26600	26600	26900	26800	26600	26400	25800	26000	26300	26600
France	24200	24500	24800	25000	25100	25300	25200	24700	24700	24700
United Kingdom	20600	20500	20300	20300	20400	20700	21000	21400	21700	22200
Poland	18200	18100	18000	17900	17800	18000	17900	17300	17100	17000
Vietnam	18100	18400	18900	19300	19800	20400	21000	21700	22400	23200
Spain	17700	17800	17800	17700	17700	17700	17300	17000	16900	16800
Ukraine	15000	14700	12800	12100	11800	12100	12000	11600	11700	11900
Turkey	15000	14600	16100	16500	16900	16900	17500	17500	17600	17900
Brazil	14800	15000	15200	15400	15700	15900	16300	16700	17200	17700

Prediction

Next Prediction the data up to three more years' i.e 2020, 2025 and 2030, how to fluctuated the trend directions on countries in this year's whether its increasing order or decreasing order. In this case Linear Regression Model is the best prediction model to represent the data. Simple linear regression is a statistical method for obtaining a formula to predict values of one variable from another where there is a causal relationship between the two variables. The variables y and x are those whose relationship we are studying.

Consider the Model function

$$y = \alpha + \beta x$$

which describes a line with slope β and y-intercept α. In general such a relationship may not hold exactly for the largely unobserved population of values of the independent and dependent variables; we call the unobserved deviations from the above equation the errors. Suppose we observe n data pairs and call them $\{(x_i, y_i), i = 1, ..., n\}$. We can describe the underlying relationship between y_i and x_i involving this error term ε_i by

$$y_i = \alpha + \beta x_i + \varepsilon_i$$

This relationship between the true (but unobserved) underlying parameters α and β and the data points is called a linear regression model.
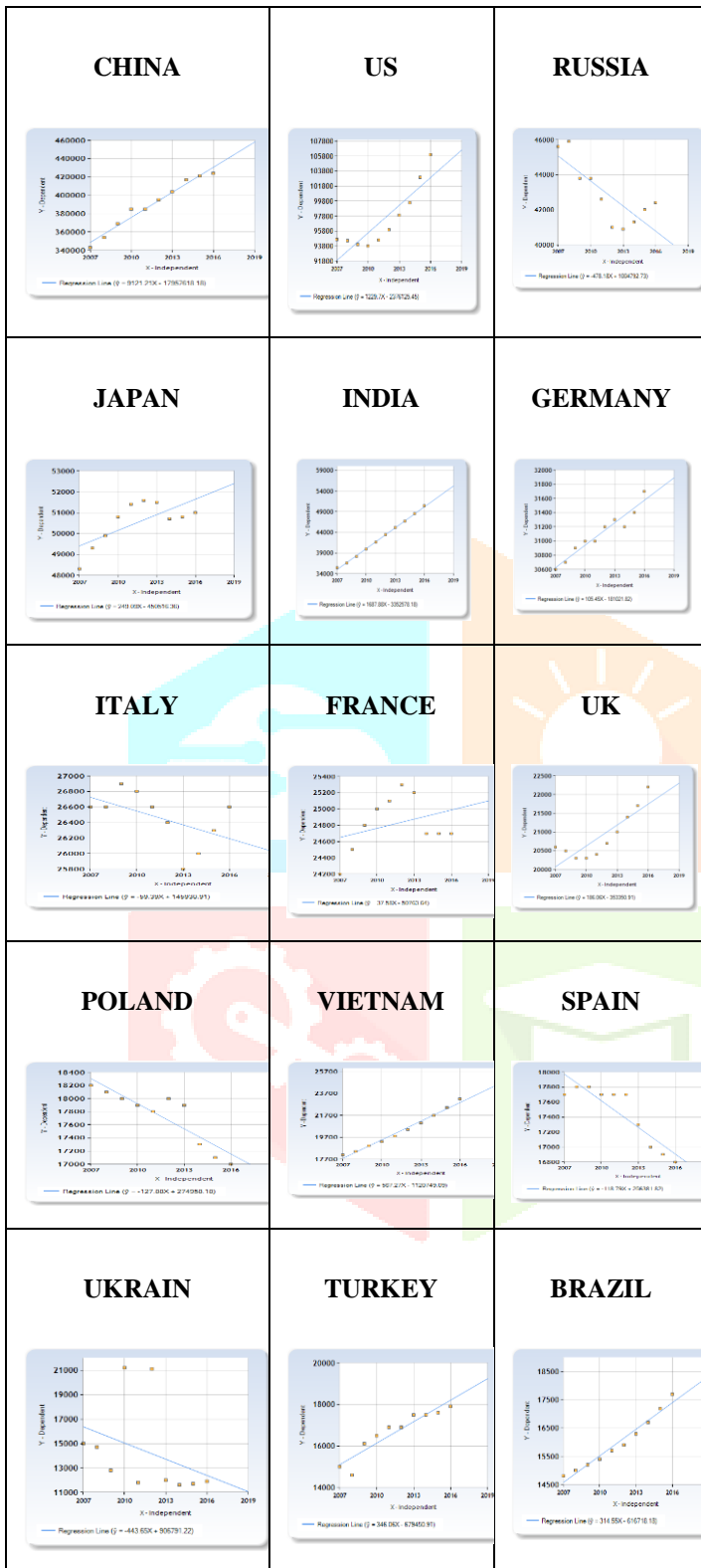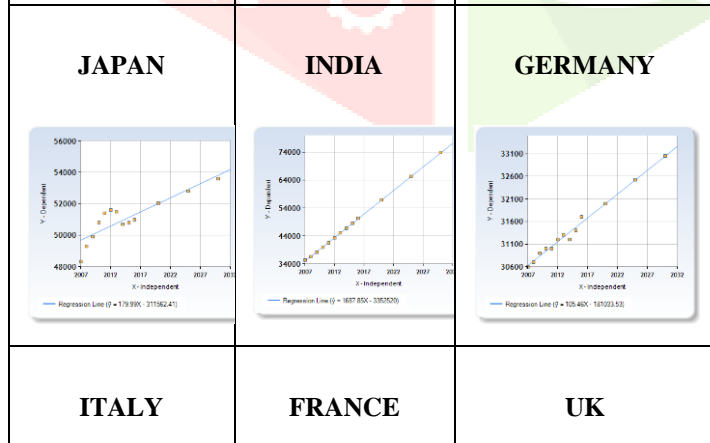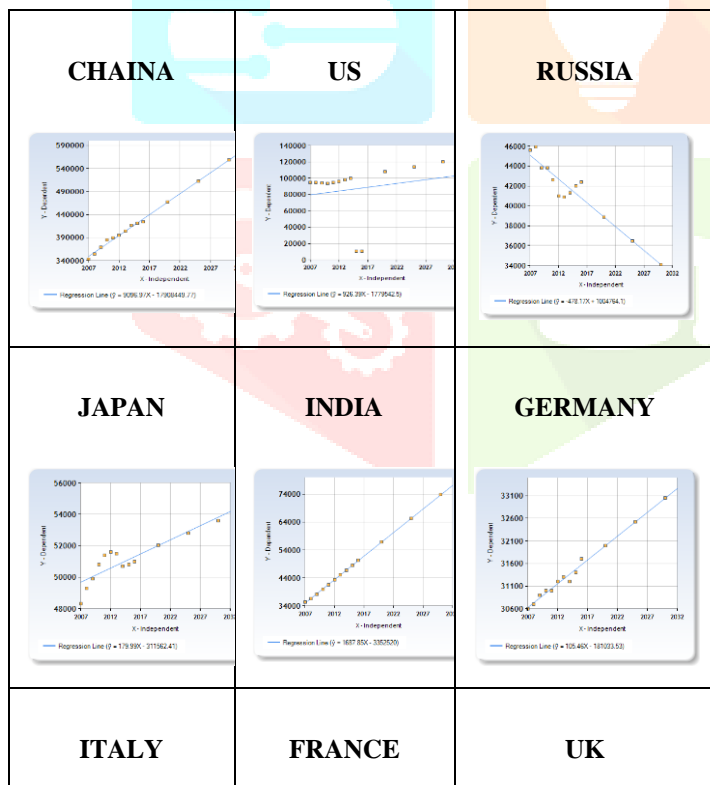
Table - 2

Country	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2020	2025	2030
China	343000	354000	369000	385000	389000	395000	404000	417000	421000	424000	467424.242	512909.0909	558393.9
United States	94700	94500	94000	93800	94600	96000	97900	99600	103000	106000	107862.424	114010.9091	120159.394
Russia	45600	45900	43800	43800	42600	41000	40900	41300	42000	42400	38865.4545	36474.5455	34083.6364
Japan	48300	49300	49900	50800	51400	51600	51500	50700	50800	51000	52031.1111	52814.4444	53597.7778
India	35400	36600	38200	39900	41600	43400	45200	46700	48500	50400	56936.9697	65376.3636	73815.7576
Germany	30600	30700	30900	31000	31000	31200	31300	31200	31400	31700	31996.3636	32523.6364	33050.9091
Italy	26600	26600	26900	26800	26600	26400	25800	26000	26300	26600	25955.1515	25658.1818	25361.2121
France	24200	24500	24800	25000	25100	25300	25200	24700	24700	24700	25139.3939	25327.2727	25515.1515
United Kingdom	20600	20500	20300	20300	20400	20700	21000	21400	21700	22200	22491.5152	23421.8182	24352.1212
Poland	18200	18100	18000	17900	17800	18000	17900	17300	17100	17000	16643.0303	16003.6364	15364.2424
Vietnam	18100	18400	18900	19300	19800	20400	21000	21700	22400	23200	25141.8182	27978.1818	30814.5455
Spain	17700	17800	17800	17700	17700	17700	17300	17000	16900	16800	16430.303	15836.3636	15242.4242
Ukraine	15000	14700	12800	12100	11800	12100	12000	11600	11700	11900	9741.8182	8078.1818	6414.5455
Turkey	15000	14600	16100	16500	16900	16900	17500	17500	17600	17900	19591.5152	21321.8182	23052.1212
Brazil	14800	15000	15200	15400	15700	15900	16300	16700	17200	17700	18663.6364	20236.36	21809.0909

CHAINA · US · RUSSIA · JAPAN · INDIA · GERMANY · ITALY · FRANCE · UK · POLAND · VIETNAM · SPAIN · UKRAIN · TUKEY · BRAZIL

Results

The model we have built can be used for inference of how the different predictors influence the outcome. It is far from perfect. Their is still presence of non-linearity and non-constant variance of errors. Moreover, the outliers and leverage points should be analyzed to find a better model. In order to use it for prediction, more concrete measures should be taken for ensuring the accuracy of the model, like cross-validation. It still helps by providing good estimations of the significant relations between the

predictors and the outcome. These estimations can be used to summarize the data in a more useful and presentful way. By the Analysis we observed that the Countries like US, Italy, Poland, Spain and Ukraine are in decreasing order while the other Countries are in increasing order. In the Countries like China, Russia, Japan, India, Germany, France, UK, Vietnam, Turkey and Brazil are going to have impact on their economical, social and biological aspects compared to other Countries.

CONCLUSION

Lung cancer is the leading cause of death over the world and the only chance of cure for patients affected from this kind of cancer is surgical resection. This review has examined international trends in lung cancer epidemiology. Emerging economies and developing countries face many challenges in initiating tobacco cessation campaigns while also addressing environmental risk factors and cultural barriers. Over the past three to four decades, in contrast, industrialized nations have seen large declines in cigarette smoking and, consequently, lung cancer. However, despite knowledge gains in tumor biology that have led to targeted therapies, mortality from lung cancer remains high for most patients around the world. Future directions must include improvements in early detection and technological advances in genomics and genetics to achieve a more personalized approach to therapy and ultimately improve lung cancer survival.

References

1. Fitzmaurice C, Dicker D, Pain A, et al. The Global Burden of Cancer 2013. JAMA Oncology. 2015; 1(4): 505–527. DOI: 10.1001/jamaoncol.2015.

2. Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J and Jemal A. 2012. Global cancer statistics, CA Cancer J Clin; 2015. DOI: 10.3322/caac.21262.

3. Thun MJ, Carter BD, Feskanich D, et al. 50-year trends in smoking-related mortality in the United States. The New England Journal of Medicine. 2013; 368(4): 351–364. DOI: 10.1056/NEJMsa1211127.

4. Youlden D, Cramb S and Baade P. . 2008; The international epidemiology of lung cancer: Geographic distribution and secular trends. Journal of Thoracic Oncology 3(8): 819–831. DOI: 10.1097/JTO.0b013e31818020eb.

5. Jemal A, Center MM, DeSantis C and Ward EM. 2010, Global patterns of cancer incidence and mortality rates and trends. Cancer Epidemiology Biomarkers & Prevention.; 19(8): 1893–1907. DOI: 10.1158/1055-9965.EPI-10-0437.

6. Goss PE, Strasser-Weippl K, Lee-Bychkovsky BL, et al. 2014; Challenges to effective cancer control in China, India, and Russia. The Lancet Oncology. 15(5): 489–538. DOI: 10.1016/S1470-2045(14)70029-4.

7. Siegel R, Ma J, Zou Z and Jemal A. Cancer statistics, 2014. CA: A Cancer Journal for Clinicians. 2014; 64(1): 9–29. DOI: 10.3322/caac.21208.

8. Freedman ND, Abnet CC, Caporaso NE, et al. 2015; Impact of changing US cigarette smoking patterns on incident cancer: Risks of 20 smoking-related cancers among the women and men of the NIH-AARP cohort. Int J Epidemiol;.

9. Meza R, Meernik C, Jeon J and Cote ML. 2015; Lung cancer incidence trends by gender, race and histology in the United States, 1973–2010. PLoS One. 10(3):

e0121323   DOI: 10.1371/journal.pone.0121323.

10. Rinh QD, Nguyen PL, Leow JJ, et al. 2015; Cancer-specific mortality of Asian Americans diagnosed with cancer: A nationwide population-based assessment. J Natl Cancer Inst. 107(6). DOI: 10.1093/jnci/djv054.

11. Bach PB, Cramer LD, Warren JL and Begg CB. 1999; Racial differences in the treatment of early-stage lung cancer. New England Journal of Medicine. 341(16): 1198–1205. DOI: 10.1056/NEJM199910143411606.

12. Wisnivesky JP, McGinn T, Henschke C, Hebert P, Iannuzzi MC and Halm EA. 2005; Ethnic disparities in the treatment of stage I non–small cell lung cancer. American Journal of Respiratory and Critical Care Medicine. 171(10): 1158–1163. DOI: 10.1164/rccm.200411-1475OC.

13. Ward E, Jemal A, Cokkinides V, et al. 2004; Cancer disparities by race/ethnicity and socioeconomic status. CA Cancer J Clin. 54(2): 78–93. DOI: 10.3322/canjclin.54.2.78.

14. UK CR. Lung cancer incidence projections to 2024: Future rates and numbers of new cases in Great Britain and the UK. Cancer Stats Cancer Projections Series; 2009.

15. Lortet-Tieulent J, Renteria E, Sharp L, et al. Convergence of decreasing male and increasing female incidence rates in major tobacco-related cancers in Europe in 1988–2010. European Journal of Cancer.

16. National Institutes of Health grant 1 RO1 CA163772.