Lung Cancer Risk Prediction using Current Human Databases: Strengths, Limitations and Future Directions

Andrew Xing[1], Zhixin Tang[2] and Zhiguang Huo[2]*

[1]*Buchholz High School, Gainesville, FL 32606*

[2]*Department of Biostatistics, Colleges of Public Health & Health Professions and Medicine, University of Florida, Gainesville, FL 32610*

*Corresponding author: Zhiguang Huo, Department of Biostatistics, Colleges of Public Health & Health Professions and Medicine, University of Florida, Gainesville, Florida 32610, United States.

Abstract

Lung cancer is the leading cause of cancer related deaths in the United States of America and the whole world. Lung cancer treatment not only has a rather limited success but also imposes tremendous financial burdens. Thus, alternative strategies are urgently needed to manage lung cancer in a more patient-friendly manner. It is extremely important for the general public to understand the chronic nature of lung cancer and to be aware of the numerous risk factors contributing to lung cancer so that preventive approaches may be implemented among the general population to complement current treatment paradigm for more effective lung cancer management. The better understanding of the different risk factors will also help identify the lung cancer high-risk individuals for early preventive interventions, which may be more effective and patient-friendly in addition to the lower cost. This manuscript discusses the various risk factors to lung cancer, including well-known risk factors, potential ones, and, importantly, emerging new risk factors that are likely to have a greater influence on the younger generation. This manuscript also discusses the complex nature of lung cancer risk factors, the application of various population-based databases for their identification and their limitations. Lastly it outlines potential future directions for lung cancer risk factor evaluation and the need for their integration in identifying individuals with higher risk of lung cancer.

Keywords: Lung cancer risk factor; Chronic nature; Tobacco smoke; Risk prediction; Human databases

Abbreviations

COPD: Chronic Obstructive Pulmonary Disease; TERT: Telomerase; CYP: Cytochrome P450 enzyme; UGT: Uridine 5'-diphospho-GlucuronosylTransferase; SNP: Single Nucleotide Polymorphism; PLCO: Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial; NCI: National Cancer Institute; EAGLE: Environmental and Genetics in Lung cancer Etiology study; CPS-II: Cancer Prevention Study II; CNV: Copy Number Variation; MRI: Magnetic Resonance Imaging; CRP: C reactive protein; nAChR: nicotinic acetylcholine receptor

Introduction

Lung cancer is the most common cancer in men and the 2nd most common cancer in women worldwide with more than 2.2 million new cases in 2020 [1]. Besides its high prevalence, the five-year overall survival rate for patients with lung cancer is very low in comparison to the other major type of malignancies, barely reaching to 23% in the United States of America in 2023 [2] while such rates are even lower in many other countries and regions [3]. The rather poor clinical outcome of patients with lung cancer are largely due to the late diagnosis of lung cancers, the majority of which are at

clinical stage upon diagnosis, the limited efficacies and significant toxicities of current therapeutic treatments, the rapid acquisition of drug resistance, and the associated high rate of disease recurrence and progression. Thus, lung cancer alone resulted in nearly 1.8 million deaths worldwide in 2020 [1], being the leading cause of the cancer-related deaths among all malignancies for decades. In addition, current lung cancer clinical management is associated with intimidating financial burdens. For instance, the cost of most targeted therapies for lung cancer are $100,000 or higher in the United States of America while the cost of the recently developed immunotherapies can reach above $400,000 [4]. Many patients with lung cancer and their families thus suffer dramatically from the associated financial toxicity with significant out-of-pocket expenses and poorer financial well-being in addition to the disease itself. Such intimidating financial burdens also contribute significantly and negatively to the compromised quality of life and reduced treatment adherence [5], all of which are associated with the rather poor outcome of lung cancer management. Therefore, besides continuous efforts to search for and to develop more effective therapeutic treatments for lung cancer, which has been the central theme for decades with significant investments and prohibitory financial burdens, paradigm-shift strategies need to be developed and implemented to improve lung cancer management, which needs to be more patient friendly-and cost-effective.

It is therefore very important to emphasize and disseminate the knowledge that lung cancer is a chronic disease; it typically takes several decades for lung cancer to evolve from initiation into the clinically detectable stage, during which minimal, if any, intervention is implemented in our current lung cancer management paradigm. Its chronic nature offers the great opportunity for early detection and early preventive intervention, which are going to be much more patient-friendly and in the long-run more cost effective, analogous to our successful managements of many other chronic diseases, such as diabetes and cardiovascular conditions. While several early diagnostic tools have been developed with certain levels of implementation in the clinic for lung cancer detection, they are mostly designed to detect lungs already with precancerous lesions or even early-stage cancers, which should have been classified as the late stage of lung cancer considering its decade-long evolving process. With the present practice and diagnostic tools, we thus still miss significant opportunities for real early detection of lung cancer and preventive intervention. To achieve more efficient and accurate early detection, comprehensive and quantitative understanding of the potential risk factors for lung cancer is required – what are the risk factors for lung cancer, what are the underlying mechanisms, what are the potential surrogate readouts, what are their potential interactions and relationships, and ideally what are their quantitative contributions to lung cancer development from a population and individual point of view. The foundational building blocks to the answers of these questions would require the prospective collection of the longitudinal data among a large number of participants given the chronic nature of lung cancer

development, the intrinsic complexity of lung carcinogenesis, the high levels of heterogeneity among human individuals, and the intrinsic random nature of genetic mutations at least based on our current knowledge. Appropriate modeling, likely artificial intelligence-driven approaches due to its complexity, is expected to be essential to efficiently analyze and interpret these longitudinal data with the ultimate goal of risk factor integration for more accurate lung cancer risk prediction. Such a risk prediction model can be potentially embedded in our current annual check-up and integrated with our annual health information, which is expected to identify, or at least enrich, the lung cancer high-risk individuals early on followed by patient-friendly and more cost effective preventive early interventions.

To help achieve this goal, the current work will first review our current knowledge about different risk factors for human lung cancer. Their applications in lung cancer risk prediction via several human prospective cohorts will be summarized as well. We will then discuss the strengths and limitations of current approaches and outline future research directions with the ultimate goal to achieve lung cancer early detection and prevention, which is essential to improve lung cancer management.

Known Lung Cancer Risk Factors

Many risk factors have been proposed for lung cancer (Figure 1).
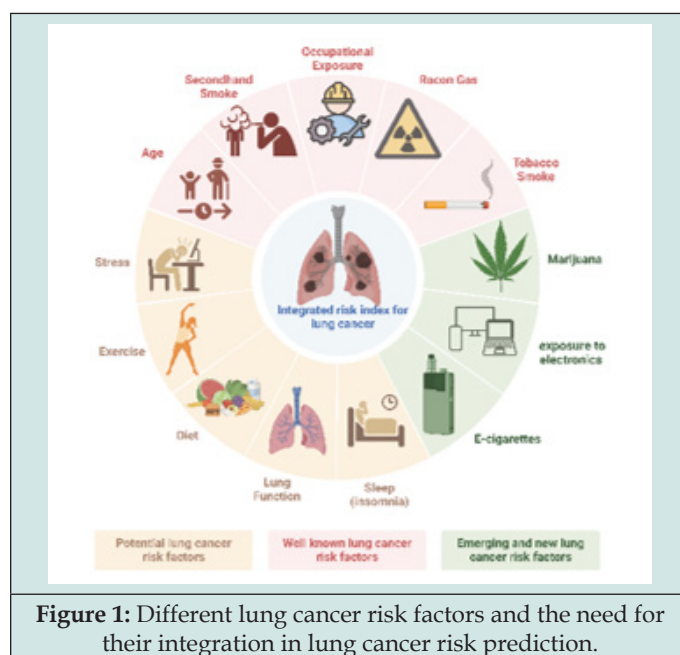

Figure 1: Different lung cancer risk factors and the need for their integration in lung cancer risk prediction.

Tobacco smoke is well-accepted as the major risk factor for lung cancer, which may have contributed to 80-90% of the lung cancer cases. With several decades of efforts, ample and compelling evidence have been accumulated, demonstrating a strong causal relationship between tobacco smoke exposure and lung cancer risk. First of all, before the start of the mass production of tobacco products in the late 19th century, lung cancer was a rare cancer [10]

although this could be partly caused by the relatively short life span of the general population during those time periods. Secondly and most importantly, a strong positive association has been observed between tobacco use and lung cancer incidence in many countries and regions with the lung cancer incidence lagging the prevalence of tobacco use by 20-30 years. This time gap between tobacco smoke exposure and lung cancer incidence also provides convincing evidence about the chronic nature of lung cancer from its initiation to the clinical stage. It should be noted that tobacco smoke is also the leading cause of numerous other preventable chronic diseases, including cardiovascular diseases, COPDs, and strokes, some of which are potentially inter-related. Despite various policies and strategies developed and implemented to control tobacco exposure, tobacco smoke is still a very serious problem in many countries, including the USA. Specifically, an estimated 46 million U.S. adults reported using tobacco products currently [11,12], resulting in an estimated half a million premature deaths each year and another 16 million living with a serious illness. Accordingly, approximately $170 billion direct medical cost is spent on smoking-related illness [11] with an overall cost of more than $800 billion [13,14] annually in the USA alone. Depressingly, around 10.0% of U.S. middle and high school students reported current (i.e., past 30-day) use of tobacco products in 2023 [15] with 22.2% of U.S. middle- and high-school students reported ever using any tobacco product, corresponding to > 6 million persons [16]. Given the addictive nature of nicotine, these middle- and high-school students are at a significant risk of continuing to use tobacco products in their lifetime, which will greatly compromise their health, including increased risk of lung cancer. In fact, the percentage of adults who smoke in the US has persisted at 13-15% since 2015 [17] and sadly there is also little indication that this number will decrease significantly in the near future [18], particularly given the prevalence of tobacco products among middle- and high-school students. Certain factors are associated with higher rate of tobacco product use, such as living in rural areas, lower education levels, and lower incomes, which are also associated with higher incidence of lung cancer. In addition, the 2020 National Youth Tobacco Survey indicates that approximately 18.4% of U.S. middle- and high-school students have used marijuana [19], which could evolve into an emerging new risk factor for lung cancer although compelling evidence will likely be not available for the next few decades, similar as the case of tobacco smoke as a lung cancer risk factor. At the same time, 10-20% of smokers eventually develop lung cancer while the rest do not, suggesting that other risk factors need to be considered in order to effectively identify or further enrich the high-risk individuals. In addition, about 10% of lung cancer patients were classified as non-smokers, further indicating the involvement of other risk factors in lung cancer development, although the "non-smokers", at least some of them, may have been passively exposed to second-hand smoking.

Besides tobacco smoke, various genetic factors have been investigated for their potential contribution to lung cancer risks, including TERT [20], CYPs [21], and UGTs [22], SNPs [23], rare germline variants [24], germline homozygosity [25], and copy number variations (CNV) [26], to name a few. The results generally showed that these genetic factors alone can only account for a very small portion of cancer risk heritability [24], even the polygenic risk models that evaluate and integrate multiple genetic risk factors [27]. These outcomes further substantiate the complex nature of lung cancer as observed in the clinics that lung cancer cases are not driven by the same genetic defects. In addition, most of these genetic analyses, if not all, have not been rigorously validated in large population-based studies [27]. Similarly, many other risk factors have been evaluated and in general their individual contribution to lung cancer risks appears to be limited or non-significant, suggesting that the integration of multiple risk factors may be essential for more effective lung cancer risk prediction, which will be discussed later.

Human cohort database and their applications in lung cancer risk factor evaluation and risk prediction realizing the complexity of lung cancer risk prediction, which requires longitudinal and comprehensive data collected from a large number of participants, several population databases have been established. The data collected, particularly the longitudinal data, have the potential to help identify lung cancer risk factors, which can be used to enrich the high-risk individuals for lung cancer. In this section we will briefly describe a few human cohort databases, analyze their strengths, identify potential limitations, and summarize their applications in human lung cancer risk prediction (Figure 2).

Figure 2: Representative human databases that have been used to identify and evaluate lung cancer risk factors.
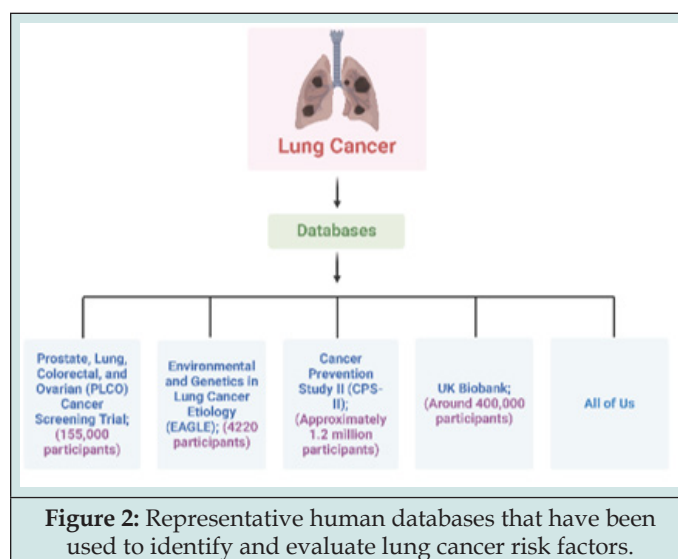
The Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial was a large randomized controlled trial designed and sponsored by the National Cancer Institute (NCI). The goal was to determine the effects of different screening methods on cancer-related mortality and secondary endpoints in men and women aged between 55 and 74. This trial enrolled approximately 155,000 participants between November 1993 and July 2001. Participants were individually randomized into the control arm or

intervention arm in equal proportions. Participants assigned to the control arm received usual care, whereas participants assigned to the intervention arm were invited to receive screening exams for prostate, lung, colorectal and ovarian cancers. Data were collected on cancer diagnoses through 2009 (median follow-up time 11.3 years) and mortality through 2018 (median-follow-up 19.2 years). All participants were asked to complete a baseline questionnaire containing information such as demographics and medical history. Intervention arm participants were also asked to complete the Dietary Questionnaire at baseline. A second dietary questionnaire was introduced in December 1998. Blood samples and buccal cell samples were also collected from certain participants for research. Around 110,000 PLCO participants were genotyped as well for genetic analyses. The cohort of PLCO participants, after following for more than two decades, resulted in a collection of lung cancer information with 1390 cases [20]. The collected information can be used for lung cancer risk factor investigation. For instance, the data have been used to develop protein-based risk biomarkers [28,29] and to evaluate the potential contributions of low-fat diet and supplements [30-33]. Orloff et al. has used the PLCO data to identify extended germline homozygosity with lung cancer risk [25]. These analyses, however, did not analyze the potential contributions to different lung cancer subtype, which is likely a major weakness of the results since different subtypes of lung cancer may have different risk factors or different contributions from the same risk factor. Indeed, Sivakumar et al. analyzed the mutation patterns between two different subtypes of lung cancer using the PLCO data and identified completely different mutation landscape [34], indicating the necessity to separate different subtypes of lung cancer in risk prediction and prevention. Environmental and genetics in lung cancer etiology study (EAGLE) is a population-based case-control study of lung cancer, including 2100 primary lung cancer cases and 2120 healthy controls enrolled in Italy between 2002 and 2005 [35] with the goal to explore the full spectrum of lung cancer etiology, from smoking addiction to lung cancer outcomes, through examination of epidemiological, molecular and clinical data. In addition to smoking data, a number of behavioral rating scales have been implemented including tobacco dependence, withdrawal, depression, anxiety, and alcohol dependence. These data have been explored for their potential to identify and evaluate different lung cancer risk factors, including gender, hormonal factors, certain gene copy numbers and microRNAs, family history, COPD, and even outdoor particulate matter [36-42]. One major limitation of this database is its nature of a single time point data collection for the enrolled participants and the rather small sample size. Thus, some results based on this database are not consistent with other studies and all of the results from this database remain to be validated in future studies.

The Cancer Prevention Study II (CPS-II), which began in 1982, is a prospective mortality study of approximately 1.2 million American men and women in all 50 states, the District of Columbia, and Puerto Rico. Each participant completed a four-page, confidential questionnaire. Baseline questions included personal identifiers, height, weight, demographic characteristics, personal and family history of cancer and other diseases, use of medicines and vitamins, menstrual and reproductive history (women), occupational exposures, dietary habits, alcohol and tobacco use, and various questions regarding exercise and behavior. Within this cohort, a CPS-II Nutrition Survey cohort was established to obtain detailed information on dietary exposures and to update with additional exposure information, and to conduct prospective cancer incidence follow-up in addition to mortality follow-up. Such new questionnaires were sent to the CPS-II Nutrition Survey cohort in 1997, 1999, 2001, 2003, 2005, 2007, 2009, 2011, 2013, and 2015. Ongoing cancer incidence follow-up for the CPS-II Nutrition Survey cohort is conducted by validating self-reported incidence cancers using medical records or linkage with state cancer registries. Nearly 30,000 incident cancers were reported in the interval 1992 to 2005, which should include over 3,000 lung cancer cases. These data could be powerful to examine the association of many surveyed factors (e.g., diet, lifestyle, and environment) with lung cancer incidence to help identify the risk factors given its longitudinal nature. Its application, however, has been limited based on the number of peer-reviewed publications; the potential reason remains unknown. Given the limitation of each individual database, particularly the limited number of lung cancer incidence, attempts to integrate data from multiple databases have been explored as well with the assumption that the major risk factors for lung cancer are similar, if not the same, among the different cohort. For instance, Landi et al. analyzed 14 databases including PLCO, EAGLE, and CPS-II, on lung cancer risk in association with different SNP [20] but failed to identify any promising candidates. Similarly, Li et al. analyzed CNV in EAGLE and PLCO on lung cancer risk without much success [26]. The negative results from these studies could be due to potential complication when integrating multiple databases. Specifically, it has been estimated that genetic factors only contribute to ~30% of lung cancer risk while environment is the major contributor; given the potential interactions between environments and genes, it may not be appropriate to combine data from different environments, including data collected from different countries and/or from the same country but during different periods of time, which are the typical variations among cohorts of populations in different database. Thus, the validity to integrate data from different database remains to be determined particularly for populations from different environments or during different period of times.

The UK biobank is a comprehensive database, collecting a wide range of data from a longitudinal cohort of general population in the United Kingdom of Britain. It contains the demographic information, biological samples (blood, saliva and urine), cognitive function, verbal interview, eye measurements, genotyped SNPs, brain MRI, cognitive function summary, mental health, work environment, local environment, diet and alcohol summary, early life experience, education and employment, genomics, geographical

and location, heart MRI, linked health outcomes, mental health, physical measurement summary, self-reported medical conditions and many other factors that may be related to human health. Within each category, various parameters have been collected as well. Using early life experience as an example, the following status have been collected – birth weight, breastfeeding status, comparative body and height size at age 10, maternal smoking around birth, part of a multiple birth, and whether being adopted or not. This prospective cohort has enrolled ~400,000 participants with periodic follow-ups to collect longitudinal data and disease outcomes. With the comprehensive data collection, numerous analyses have evaluated the predictive power for a wide range of potential lung cancer risk factors, including tobacco smoke, stressful life experience, inflammation, lung function, sleep, circadian rhythm, and many other risk factor candidates with some positive indications [43-46]. For instance, CRP, an inflammation biomarker, has been demonstrated to increase lung cancer risk, including all lung cancer subtypes in the Biobank samples [47]. The level of bilirubin in the blood appeared to reduce lung cancer risk although the subtypes of lung cancer were not differentiated [48]. Polygenetic prediction has been employed for lung cancer risk prediction as well [27], even in the context of smoking [49]. Similarly, the genetic and smoking interaction in lung cancer risk prediction has been explored via unbiased approach [50]. Sleep [51, 52] and neurological functions/stress [53, 54] as a lung cancer risk factor has been explored with interesting results. Specifically, psychological stress increased lung cancer risk among non-smokers, light smokers and heavy smokers by 43.0%, 46.8%, and 31.8% respectively [54] and a causal relationship was demonstrated as well in the same cohort [53]. This is consistent with human epidemiological data that individuals with mental health issues are at a higher risk of lung cancer based on a meta-analysis of 165 longitudinal studies [55]. Integrating multiple risk factors, including stress, smoking and genetic status, appear to result in better lung cancer risk predictions [54], however, systematic investigations have not been done with varied combinations of risk factor candidates, since some risk factors may be redundant or have interactions, such as stress and sleep. Additional risk factors evaluated using the Biobank data include walking [56], green tea [57], beta-blocker [58], diabetic status [59], asthma [60], polygenic risk factors [61], plasma protein markers [62], telomere length [63] and many others [64-67]. Recently Krishna et al utilized the Biobank and reported the association of HLA-II heterozygosity with reduced risk of lung cancer, implying that genetic variations in immune surveillance is a key feature of cancer susceptibility, together with environmental exposures [68].

Interestingly, depression and anxiety have been found to contribute to increased risk of lung cancer but no other cancers in this cohort [69]. Pettit et al also analyzed a range of heritable traits as lung cancer risk [70]. Once again, a wide range of factors have been evaluated for their potential in lung cancer risk prediction with many of them showing limited levels of predictive power, demonstrating the challenges and complexity of lung cancer risk

prediction. The limited predictive power for each individual risk factor also suggests that multiple factors need to be integrated while factors associated with tobacco smoke may be of greater contributions, such as daily tobacco exposure level, genetic factors associated with tobacco use and tobacco toxicant metabolism, such as CYP2A6, and genetic factors to nAChRs. At the same time, different risk factors may have interactions, redundancy, causal relationship, and other more complicated associations. Thus, none of these factors alone were sufficient or powerful enough to predict lung cancer risk that their integrations are likely necessary. The data in Biobank offers the great opportunity to explore these potentials, particularly given its longitudinal nature that data will become more and more comprehensive with more lung cancer cases and hopefully more powerful to support such discoveries. Similar to the Biobank at UK, All of US is another prospective cohort of population in the USA. Its application in lung cancer risk exploration has been limited at this point, potentially because of the short period of longitudinal data collection to date since this cohort was established later than the UK Biobank. However, it will offer the additional opportunity similar as Biobank with more longitudinal data collected. Besides the targeted risk factor analyses in these database application, unbiased omics techniques have also been employed to identify potential risk factors in these studies with limited success. Although there are many unique strengths of the unbiased approach, the sensitivity of these methods remains to be determined. It is also possible that no single parameter, such as SNP, is powerful enough as an independent risk factor similar to many targeted risk candidates evaluated. In addition, some of the unbiased profiling may need to be interrogated in the context of specific environmental conditions, such as smoking status since some SNPs in nicotine addiction and tobacco toxicant metabolism may be a risk factor only in the context of tobacco smoke exposure that such analyses will only be valid among the participants who smoke, not the whole population in the database.

Future Directions

The current databases also have certain limitations-some risk factors are not well documented, such as radon exposure, tobacco smoke exposure information (no information about the tobacco products used by the individuals, the limitations of survey-based qualitative information of tobacco exposure, and the lack of biological quantification), second-hand smoking, environmental, occupational and domestic pollutions. There are also emerging new risk factors given the life-style changes, such as the increased prevalence of electronics, the reduction in physical activities, the changes in diets and sleeping patterns, and many other life style changes. There are also intrinsic risks for the integration of different database, because the causes for lung cancer are evolving and potentially different for different regions during different time periods. For example, the causes for lung cancer in the USA now could be substantially different from what they were in two or three decades ago, such as tobacco use, tobacco products, use

of electronics, change in physical activities, and many other life style change. The causes for different subtypes of lung cancer can be different too: although tobacco smoke is the main cause of lung cancer, other factors may be involved for the different subtypes of lung cancer. Thus, if possible, different subtypes of lung cancer should be studied separately. The causes for lung cancer among different populations could be different as well although the difference could be subtle: there have been ample data suggesting lung cancer risk disparity with respect to race, gender, and other factors. Some of these may be driven by genetic factors and some may be driven by environmental factors.

In summary, with the continuous growth of these large prospective cohort databases, such as the UK Biobank and All of US, their longitudinal data collection, more comprehensive data on different risk factors, and the integration of multiple risk factors, these databases are expected to become more powerful in identify individual lung cancer risk factors, quantifying their potential contributions, and more importantly developing integrated risk index for more accurate lung cancer risk prediction. Given the complexity of lung cancer risks, artificial intelligence may be essential to help analyze the different risk factors, explore their potential interactions, and holistically integrate them for better risk prediction.

Acknowledgement

Conflict of Interest

No conflict of interest.

References

1. Huang J, Yunyang Deng, Man Sing Tin, Veeleah Lok, Chun Ho Ngai, et al., (2022) Distribution, Risk Factors, and Temporal Trends for Lung Cancer Incidence and Mortality: A Global Analysis. Chest. 161(4): 1101-1111.

2. Siegel RL, Miller KD, Wagle NS, Jemal A (2023) Cancer statistics. CA Cancer J Clin. 73(1): 17-48.

3. Sung H, Jacques Ferlay, Rebecca L Siegel, Mathieu Laversanne, Isabelle Soerjomataram, et al., (2021) Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin. 71(3): 209-249.

4. Schaft N, Jan Dörrie, Gerold Schuler, Beatrice Schuler-Thurner, Husam Sallam, et al., (2023) The future of affordable cancer immunotherapy. Front Immunol. 14: 1248867.

5. Tran G, Zafar SY (2018) Financial toxicity and implications for cancer care in the era of molecular and immune therapies. Ann Transl Med. 6(9): 166.

6. Thandra KC, Adam Barsouk, Kalyan Saginala, John Sukumar Aluru, Alexander Barsouk, et al., (2021) Epidemiology of lung cancer. Contemp Oncol (Pozn). 25(1): 45-52.

7. Gu F, Sholom Wacholder, Stephanie Kovalchik, Orestis A Panagiotou, Carolyn Reyes-Guzman, et al., (2014)Time to smoke first morning cigarette and lung cancer in a case-control study. J Natl Cancer Inst. 106(6): dju118.

8. Barta JA, CA Powell, JP Wisnivesky (2019) Global Epidemiology of Lung Cancer. Ann Glob Health. 85(1): 8.

9. Leiter A, RR Veluswamy, JP Wisnivesky (2023) The global burden of lung cancer: current status and future trends. Nat Rev Clin Oncol. 20(9): 624-639.

10. Islami F, LA Torre, A Jemal (2015) Global trends of lung cancer mortality and smoking prevalence. Transl Lung Cancer Res. 4(4): 327-338.

11. (2018) Prevention C.f.D.C.a. Centers for Disease Control and Prevention. [cited 2018 09/22/2018].

12. Wang TW, Kat Asman, Andrea S Gentzke, Karen A Cullen, Enver Holder-Hayes, et al., (2018) Tobacco Product Use Among Adults - United States, 2017. MMWR Morb Mortal Wkly Rep. 67(44): 1225-1232.

13. Shrestha SS, Ramesh Ghimire, Xu Wang, Katrina F Trivers, David M Homa, et al., (2022) Cost of Cigarette Smoking–Attributable Productivity Losses, U.S., 2018. Am J Prev Med. 63(4): 478-485.

14. Max W, HY Sung, Y Shi (2012) Deaths from second-hand smoke exposure in the United States: economic implications. Am J Public Health. 102(11): 2173-2180.

15. Jebai R, Olatokunbo Osibogun, Wei Li, Prem Gautam, Zoran Bursac, et al., (2023) Temporal Trends in Tobacco Product Use Among US Middle and High School Students: National Youth Tobacco Survey, 2011-2020. Public Health Rep. 138(3): 483-492.

16. Birdsey J, Monica Cornelius, Ahmed Jamal, Eunice Park-Lee, Maria R Cooper, et al., (2023) Tobacco Product Use Among U.S. Middle and High School Students - National Youth Tobacco Survey, 2023. MMWR Morb Mortal Wkly Rep. 72(44): 1173-1182.

17. Cornelius ME, Caitlin G Loretan, Ahmed Jamal, Brittny C Davis Lynn, Margaret Mayer, et al., (2023) Tobacco Product Use Among Adults - United States, 2021. MMWR Morb Mortal Wkly Rep. 72(18): 475-483.

18. (2004) Cancer I.A.f.R.o., Tobacco smoke and involuntary smoking. IARC monographs on the evaluation of carcinogenic risks to human. RARC: 53-119.

19. Rao S, CJ Peterson, S Yang, K Nugent (2023) Marijuana Use in Middle and High School Students: Insights from the 2020 National Youth Tobacco Survey. South Med J. 116(3): 279-285.

20. Landi MT, Nilanjan Chatterjee, Kai Yu, Lynn R Goldin, Alisa M Goldstein, et al., (2011) A Genome-wide Association Study of Lung Cancer Identifies a Region of Chromosome 5p15 Associated with Risk for Adenocarcinoma. Am J Hum Genet. 88(6): 861.

21. Timofeeva MN, Silke Kropp, Wiebke Sauter, Lars Beckmann, Albert Rosenberger, et al., (2009) CYP450 polymorphisms as risk factors for early-onset lung cancer: gender-specific differences. Carcinogenesis. 30(7): 1161-1169.

22. Gallagher CJ, Joshua E Muscat, Amy N Hicks, Yan Zheng, Anne-Marie Dyer, et al., (2007) The UDP-glucuronosyltransferase 2B17 gene deletion polymorphism: sex-specific association with urinary 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanol glucuronidation phenotype and risk for lung cancer. Cancer Epidemiol Biomarkers Prev. 16(4): 823-828.

23. Li X, Qijun Wu , Baosen Zhou, Yashu Liu, Jiale Lv, et al., (2021) Umbrella Review on Associations Between Single Nucleotide Polymorphisms and Lung Cancer Risk. Front Mol Biosci. 8: 687105.

24. Sang J, Tongwu Zhang, Jung Kim, Mengying Li, Angela C Pesatori, et al., (2022) Rare germline deleterious variants increase susceptibility for lung cancer. Hum Mol Genet. 31(20): 3558-3565.

25. Orloff MS, L Zhang, G Bebek, C Engn(2012) Integrative genomic analysis reveals extended germline homozygosity with lung cancer risk in the PLCO cohort. PLoS One. 7(2): e31975.

26. Li X, Xianfeng Chen, Guohong Hu, Yang Liu, Zhenguo Zhang, et al., (2014) Combined analysis with copy number variation identifies risk loci in lung cancer. Biomed Res Int. 2014: 469103.

27. Hung RJ, Matthew T Warkentin, Yonathan Brhane, Nilanjan Chatterjee, David C Christiani, et al., (2021) Assessing Lung Cancer Absolute Risk Trajectory Based on a Polygenic Risk Model. Cancer Res. 81(6): 1607-1615.

28. Irajizad E, Johannes F Fahrmann, Tracey Marsh, Jody Vykoukal, Jennifer B Dennison, et al., (2023) Mortality Benefit of a Blood-Based Biomarker Panel for Lung Cancer on the Basis of the Prostate, Lung, Colorectal, and Ovarian Cohort. J Clin Oncol. 41(27): 4360-4368.

29. Fahrmann JF, Tracey Marsh, Ehsan Irajizad, Nikul Patel, Eunice Murage, et al., (2022) Blood-Based Biomarker Panel for Personalized Lung Cancer Risk Assessment. J Clin Oncol. 40(8): 876-883.

30. Zhu Z, Linglong Peng, He Zhou, Haitao Gu, Yunhao Tang, et al., (2023) Low-fat dairy consumption and the risk of lung cancer: A large prospective cohort study. Cancer Med. 12(15): 16558-16569.

31. Zhu Z, Linglong Peng, Haitao Gu, Yunhao Tang, Yi Xiao, et al., (2023) Association between dietary approaches to stop hypertension eating pattern and lung cancer risk in 98,459 participants: results from a large perspective study. Front Nut. 10: 1142067.

32. Zhang Y, Guochao Zhong, Min Zhu, Ling Chen, Huajing Wan, et al., (2022) Association Between Diabetes Risk Reduction Diet and Lung Cancer Risk in 98,159 Participants: Results from a Prospective Study. Front Oncol. 12: 855101.

33. Wang Q, Meng Ru, Yaning Zhang, Tamara Kurbanova, Paolo Boffetta (2021) Dietary phytoestrogen intake and lung cancer risk: an analysis of the Prostate, Lung, Colorectal and Ovarian (PLCO) cancer screening trial. Carcinogenesis. 42(10): 1250-1259.

34. Sivakumar S, F Anthony San Lucas, Tina L McDowell, Wenhua Lang, Li Xu, et al., (2017) Genomic Landscape of Atypical Adenomatous Hyperplasia Reveals Divergent Modes to Lung Adenocarcinoma. Cancer Res. 77(22): 6119-6130.

35. Landi MT, Dario Consonni, Melissa Rotunno, Andrew W Bergen, Alisa M Goldstein, et al., (2008) Environment and Genetics in Lung cancer Etiology (EAGLE) study: an integrative population-based case-control study of lung cancer. BMC Public Health. 8: 203.

36. Pesatori AC, Michele Carugno, Dario Consonni, Neil E Caporaso, Sholom Wacholder, et al., (2013) Reproductive and hormonal factors and the risk of lung cancer: the EAGLE study. Int J Cancer. 132(11): 2630-2639.

37. Landi MT, Yingdong Zhao, Melissa Rotunno, Jill Koshiol, Hui Liu, et al., (2010) MicroRNA expression differentiates histology and predicts survival of lung cancer. Clin Cancer Res. 16(2): 430-441.

38. Rotunno M, Tram K Lam, Aurelie Vogt, Pier Alberto Bertazzi, Jay H Lubin, et al., (2012) GSTM1 and GSTT1 copy numbers and mRNA expression in lung cancer. Mol Carcinog. 51 Suppl 1(Suppl 1): E142-150.

39. Gao Y, Alisa M Goldstein, Dario Consonni, Angela C Pesatori, Sholom Wacholder, et al., (2009) Family history of cancer and non-malignant lung diseases as risk factors for lung cancer. Int J Cancer. 125(1): 146-152.

40. Koshiol J, Melissa Rotunno, Dario Consonni, Angela Cecilia Pesatori, Sara De Matteis, et al., (2009) Chronic obstructive pulmonary disease and altered risk of lung cancer in a population-based case-control study. PLoS One. 4(10): e7380.

41. Consonni D, Michele Carugno, Sara De Matteis, Francesco Nordio, Giorgia Randi, et al., (2018) Outdoor particulate matter (PM10) exposure and lung cancer risk in the EAGLE study. PLoS One. 13(9): e0203539.

42. De Matteis S, Dario Consonni, Angela C Pesatori, Andrew W Bergen, Pier Alberto Bertazzi, et al., (2013) Are women who smoke at higher risk for lung cancer than men who smoke? Am J Epidemiol. 177(7): 601-612.

43. Baumeister SE, Hansjörg Baurecht, Michael Nolde, Zoheir Alayash, Sven Gläser, et al., (2021) Cannabis Use, Pulmonary Function, and Lung Cancer Susceptibility: A Mendelian Randomization Study. J Thorac Oncol. 16(7): 1127-1135.

44. Kachuri L, Mattias Johansson, Sara R Rashkin, Rebecca E Graff, Yohan Bossé, et al., (2020) Immune-mediated genetic pathways resulting in pulmonary function impairment increase lung cancer susceptibility. Nat Commun. 11(1): 27.

45. Warkentin MT, S Lam, RJ Hung (2019) Determinants of impaired lung function and lung cancer prediction among never-smokers in the UK Biobank cohort. EBioMedicine. 47: 58-64.

46. Muller DC, M Johansson, Brennan (2017) Lung Cancer Risk Prediction Model Incorporating Lung Function: Development and Validation in the UK Biobank Prospective Cohort Study. J Clin Oncol. 35(8): 861-869.

47. Ji M, Lingbin Du, Zhimin Ma, Junxing Xie, Yanqian Huang, et al., (2022) Circulating C-reactive protein increases lung cancer risk: Results from a prospective cohort of UK Biobank. Int J Cancer. 150(1): 47-55.

48. Horsfall LJ, S Burgess, I Hall, I Nazareth (2020) genetically raised serum bilirubin levels and lung cancer: a cohort study and Mendelian randomisation using UK Biobank. Thorax. 75(11): 955-964.

49. Zhang P, Pei-Liang Chen, Zhi-Hao Li, Ao Zhang, Xi-Ru Zhang, et al., (2022) Association of smoking and polygenic risk with the incidence of lung cancer: a prospective cohort study. Br J Cancer. 126(11): 1637-1646.

50. Jia G, Wanqing Wen, Pierre P Massion, Xiao-Ou Shu, Wei Zheng, et al., (2021) Incorporating both genetic and tobacco smoking data to identify high-risk smokers for lung cancer screening. Carcinogenesis. 42(6): 874-879.

51. Peeri NC, MH Tao, S Demissie, UDT Nguyen (2022) Sleep Duration, Chronotype, and Insomnia and the Risk of Lung Cancer: United Kingdom Biobank Cohort. Cancer Epidemiol Biomarkers Prev. 31(4): 766-774.

52. Xie J, Meng Zhu, Mengmeng Ji, Jingyi Fan, Yanqian Huang, et al., (2021) Relationships between sleep traits and lung cancer risk: a prospective cohort study in UK Biobank. Sleep. 44(9).

53. Wei X, Xiangxiang Jiang, Xu hang, Xikang Fan, Mengmeng Ji, et al., (2022) Association Between Neuroticism and Risk of Lung Cancer: Results from Observational and Mendelian Randomization Analyses. Front Oncol. 12: 836159.

54. Zhang J, Yi Wang, Tingting Hua, Xiaoxia Wei, Xiangxiang Jiang, et al., (2023) Association of psychological distress, smoking and genetic risk with the incidence of lung cancer: a large prospective population-based cohort study. Front Oncol. 13: 1133668.

55. Chida Y, M Hamer, J Wardle, A Steptoe (2008) Do stress-related psychosocial factors contribute to cancer incidence and survival? Nat Clin Pract Oncol. 5(8): 466-475.

56. Chen F, Chutong Lin, Xing Gu, Yingze Ning, Huayu He, et al., (2024) Exploring the link between walking and lung cancer risk: a two-stage Mendelian randomization analysis. BMC Pulm Med. 24(1): 129.

57. Lu J, Ye Lin, Junfei Jiang, Lei Gao, Zhimin Shen, et al., (2024) Investigating the potential causal association between consumption of green tea and risk of lung cancer: a study utilizing Mendelian randomization. Front Nutr. 11: 1265878.

58. Lu Y, Jiachun Luo, Zhenyu Huo, Fan Ge, Yang Chen, et al., (2023) Causal effect of beta-blockers on the risk of lung cancer: a Mendelian randomization study. J Thorac Dis. 15(12): 6651-6660.

59. Hua J, Huan Lin, Xiaojie Wang, Zhengmin Min Qian, Michael G Vaughn, et al., (2024) Associations of glycosylated hemoglobin, pre-diabetes, and type 2 diabetes with incident lung cancer: A large prospective cohort study. Diabetes Metab Syndr. 18(2): 102968.

60. Huang Q, Yunxia Huang, Senkai Xu, Xiaojun Yuan, Xinqi Liu, et al., (2024) Association of asthma and lung cancer risk: A pool of cohort studies and Mendelian randomization analysis. Medicine (Baltimore). 103(5): e35060.

61. Hu J, Y Ye, G Zhou, H Zhao (2024) Using clinical and genetic risk factors for risk prediction of 8 cancers in the UK Biobank. JNCI Cancer Spectr. 8(2).

62. Li H, Sha Du, Jinglan Dai, Yunke Jiang, Zaiming Li, et al., (2024) Proteome-wide Mendelian randomization identifies causal plasma proteins in lung cancer. iScience. 27(2): 108985.

63. Wong JY, Batel Blechter, Aubrey K Hubbard, Mitchell J Machiela, Jianxin Shi, et al., (2024) Phenotypic and genetically predicted leucocyte telomere length and lung cancer risk in the prospective UK Biobank. Thorax. 79(3): 274-278.

64. Christakoudi S, KK Tsilidis, E Evangelou, E Riboli (2023) Interactions of platelets with obesity in relation to lung cancer risk in the UK Biobank cohort. Respir Res. 24(1): 249.

65. Zhang S, Lei Liu, Shanshan Shi, Heng He, Qian Shen, et al., (2024) Bidirectional Association Between Cardiovascular Disease and Lung Cancer in a Prospective Cohort Study. J Thorac Oncol. 19(1): 80-93.

66. Li M, Su-Mei Cao, Niki Dimou, Lan Wu, Ji-Bin Li, et al., (2024) Association of Metabolic Syndrome With Risk of Lung Cancer: A Population-Based Prospective Cohort Study. Chest. 165(1): 213-223.

67. He H, Ming-Ming He, Haoxue Wang, Weihong Qiu, Lei Liu, et al., (2023) In Utero and Childhood/Adolescence Exposure to Tobacco Smoke, Genetic Risk, and Lung Cancer Incidence and Mortality in Adulthood. Am J Respir Crit Care Med. 207(2): 173-182.

68. Krishna C, Anniina Tervi, Miriam Saffern, Eric A Wilson, Seong-Keun Yoo, et al., (2024) An immunogenetic basis for lung cancer risk. Science. 383(6685): eadi3808.

69. van Tuijl, LA, Maartje Basten, Kuan-Yu Pan, Roel Vermeulen, Lützen Portengen, et al., (2023) Depression, anxiety, and the risk of cancer: An individual participant data meta-analysis. Cancer. 129(20): 3287-3299.

70. Pettit RW, Jinyoung Byun, Younghun Han, Quinn T Ostrom, Cristian Coarfa, et al., (2023) Heritable Traits and Lung Cancer Risk: A Two-Sample Mendelian Randomization Study. Cancer Epidemiol Biomarkers Prev. 32(10): 1421-1435.

To Submit Your Article Click Here: **Submit Article**

Lupine Online Journal of Medical Sciences

Assets of Publishing with us

- Global archiving of articles
- Immediate, unrestricted online access
- Rigorous Peer Review Process
- Authors Retain Copyrights
- Unique DOI for all articles