

RESEARCH

Open Access



Construction and case study of a novel lung cancer risk index

Ali Faghani^{1†}, Lei Guo^{2†} , Margaret E. Wright³ , M. Courtney Hughes⁴  and Mahdi Vaezi^{1*} 

Abstract

Purpose: This study constructs a lung cancer risk index (LCRI) that incorporates many modifiable risk factors using an easily reproducible and adaptable method that relies on publicly available data.

Methods: We used meta-analysis followed by Analytic Hierarchy Process (AHP) to generate a lung cancer risk index (LCRI) that incorporates seven modifiable risk factors (active smoking, indoor air pollution, occupational exposure, alcohol consumption, secondhand smoke exposure, outdoor air pollution, and radon exposure) for lung cancer. Using county-level population data, we then performed a case study in which we tailored the LCRI for use in the state of Illinois (LCRI_{IL}).

Results: For both the LCRI and the LCRI_{IL}, active smoking had the highest weights (46.1% and 70%, respectively), whereas radon had the lowest weights (3.0% and 5.7%, respectively). The weights for alcohol consumption were 7.8% and 14.7% for the LCRI and the LCRI_{IL}, respectively, and were 3.8% and 0.95% for outdoor air pollution. Three variables were only included in the LCRI: indoor air pollution (18.5%), occupational exposure (13.2%), and secondhand smoke exposure (7.6%). The Consistency Ratio (CR) was well below the 0.1 cut point. The LCRI_{IL} was moderate though significantly correlated with age-adjusted lung cancer incidence ($r = 0.449$, $P < 0.05$) and mortality rates ($r = 0.495$, $P < 0.05$).

Conclusion: This study presents an index that incorporates multiple modifiable risk factors for lung cancer into one composite score. Since the LCRI allows data comprising the composite score to vary based on the location of interest, this measurement tool can be used for any geographic location where population-based data for individual risk factors exist. Researchers, policymakers, and public health professionals may utilize this framework to determine areas that are most in need of lung cancer-related interventions and resources.

Keywords: Lung cancer, Risk factors, Analytic hierarchy processes, Meta-analysis, Risk index

Introduction

Cancer is the second leading cause of death in the US, with lung cancer accounting for almost one-quarter of these deaths. The American Cancer Society estimates that 236,740 new lung cancers will be diagnosed in 2022, and this disease will claim the lives of more than 130,000

men and women [1]. Numerous studies have examined risk factors for lung cancer, with smoking being the single largest contributor to the disease [2–11]. Other established risk factors include age [12], secondhand smoke exposure [13], environmental exposures (radon [14], indoor and outdoor air pollution [15, 16]), occupational exposures [17], diet [18], alcohol consumption [19], genetic predisposition [20], previous lung disease [21], and arsenic exposure [22]. Many of these risk factors are modifiable, including active smoking and secondhand smoke exposure, environmental exposures, occupational exposures, alcohol consumption, and diet [23].

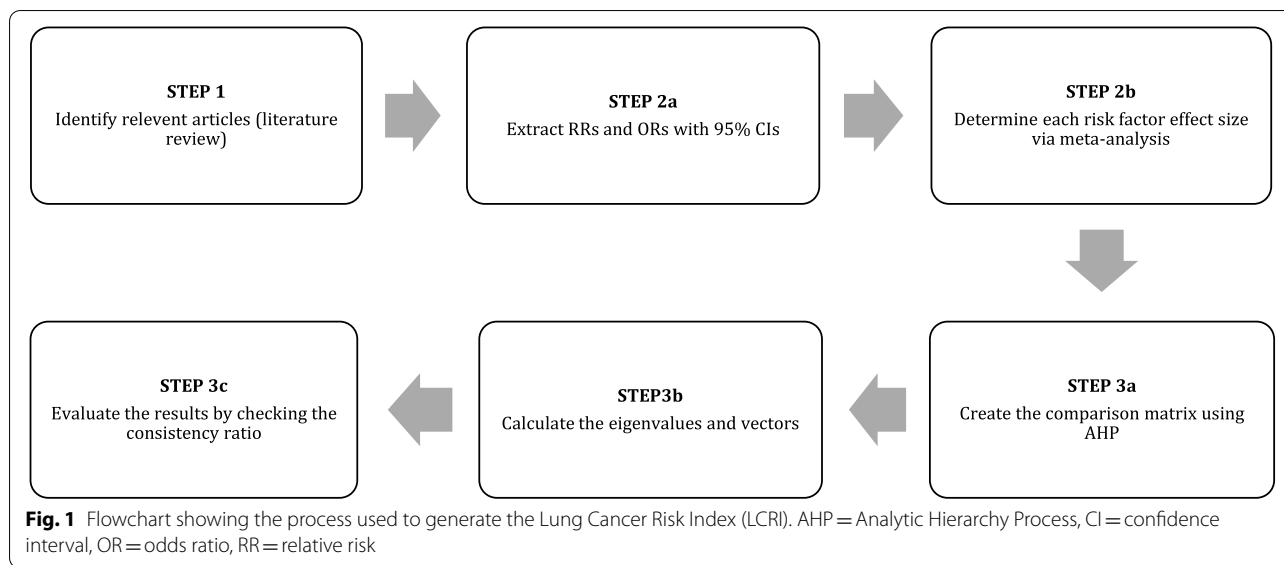
[†]Ali Faghani and Lei Guo contributed equally to this work.

*Correspondence: mvaezi@niu.edu

¹ College of Engineering and Engineering Technology, Northern Illinois University, DeKalb, IL, USA

Full list of author information is available at the end of the article





Although many studies have investigated associations between individual risk factors and lung cancer risk or mortality [20–32], less is known about how these factors interact to influence the development and progression of the disease. Some studies have examined interactions between smoking and one other risk factor, such as radon, alcohol consumption, family history, previous lung disease, or some component of diet [33]. To our knowledge, there are few, if any, studies that simultaneously investigated the contribution of more than two modifiable risk factors for lung cancer. This may be because epidemiologic studies are often limited in their ability to consider multiple factors simultaneously, given limited sample sizes and ranges of exposures within their study populations [34].

To address this gap, we constructed a Lung Cancer Risk Index (LCRI) that incorporates several modifiable risk factors using Meta-Analytic Hierarchy Process (Meta-AHP). While this approach has been used in the soil science field [35], it has not been commonly employed in the health sciences. Meta-AHP may be superior to a traditional principal component analysis approach because Meta-AHP can effectively extract essential variables and assign weights more precisely. We tailored this index for use in a case study of the state of Illinois; the $LCRI_{IL}$ was created using publicly available county-level data for all 102 Illinois counties. We then evaluated the correlation between the $LCRI_{IL}$ and reported lung cancer incidence and mortality rates. We provide researchers with an easily reproducible and adaptable method that uses publicly available data to generate a composite measure that integrates multiple modifiable risk factors for lung cancer. This measure can be tailored for any geographic area

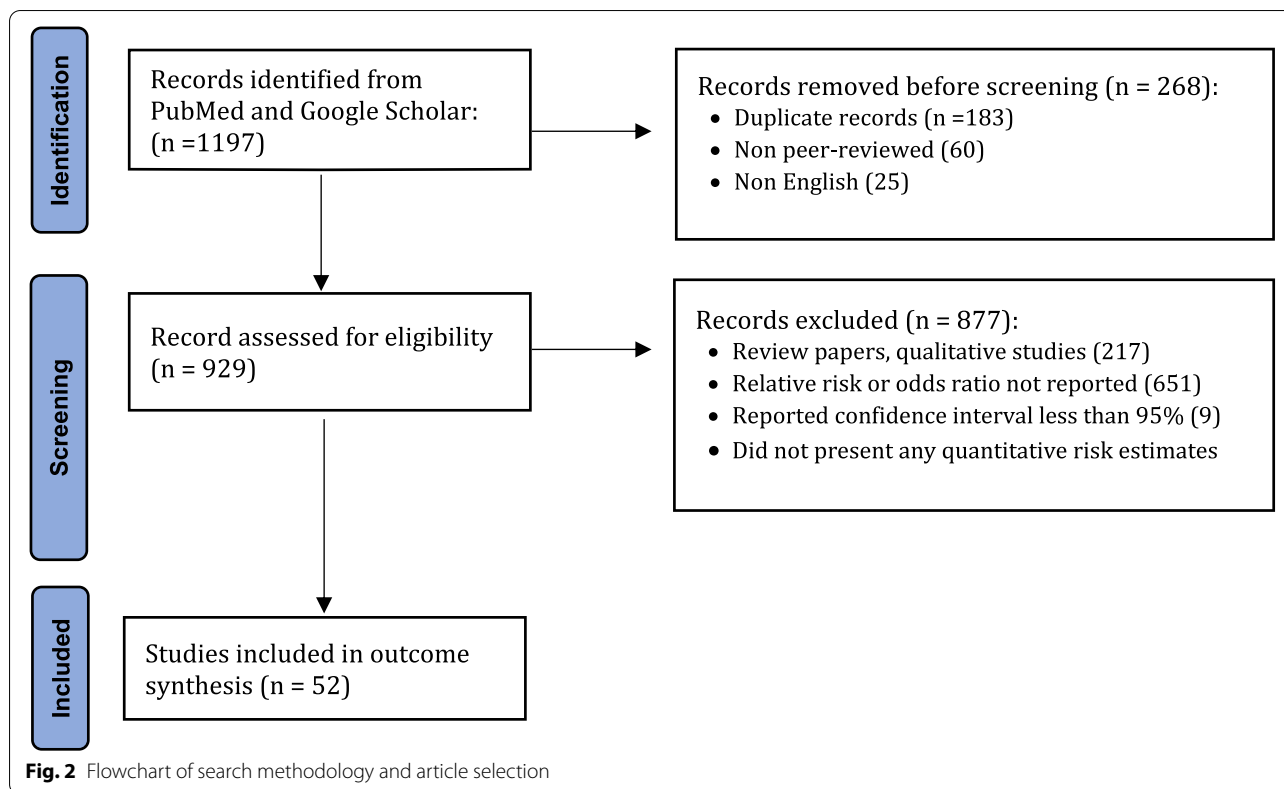
and is potentially widely applicable. Public health officials and policymakers may consider using this measure when making decisions regarding lung cancer-related interventions and resource allocation in their communities.

Methods

Figure 1 shows the process that we used to generate the lung cancer risk index (LCRI). Each step in the figure is explained in detail below.

Step 1: identify relevant articles: search strategy and article selection

Using the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines [36], we conducted searches of PubMed (including MEDLINE) and Google Scholar for full-length articles that were published between January 1990 and April 2021. We utilized the following keyword strings to capture relevant studies: “lung cancer” in conjunction with one of the following—“smoking,” “passive smoking,” “secondhand smoke,” “radon,” “occupation,” “air pollution,” “alcohol consumption,” or “risk factors.” We did not include diet in our index because the World Cancer Research Fund (WCRF) and American Institute for Cancer Research (AICR) consider there to be “limited evidence” that diet is a risk factor for lung cancer [37]. We chose to exclude arsenic exposure from our index because the US public water supply levels are kept below 50 $\mu\text{g/L}$ [38, 39], which is far below concentrations associated with increased lung cancer risk [22, 40]. Nevertheless, researchers in other countries should consider adding arsenic to an LCRI adapted for use in their locations. We assessed the quality of the articles included in the present study using appraisal



checklists and criteria of quality recommended by JBI (formerly known as "Joanna Briggs Institute"), an international organization focused on improving evidence as it relates to the feasibility, appropriateness, meaningfulness, and effectiveness of healthcare interventions [41].

As shown in Fig. 2, the initial literature search yielded 1197 articles. We removed 268 articles that were duplicates, not peer-reviewed prior to publication, or written in languages other than English. We then reviewed the abstracts of the 929 remaining articles and applied the study inclusion criteria: (1) randomized controlled trial, prospective cohort study, retrospective cohort study, case-cohort study, case-control study, or nested case-control study; (2) reported the relative risk (RR) or odds ratio (OR) associated with *increased* risk (i.e., RR or OR > 1, which is a requirement of the Analytic Hierarchy Process (AHP) model); and (3) reported 95% confidence intervals (CIs). After excluding 877 articles that did not meet the inclusion criteria specified above, at least two researchers reviewed the full text of the remaining 52 manuscripts [42].

Steps 2a / 2b: meta-analysis

The second step in creating our index was to extract the adjusted OR and RR from all 52 articles for each lung cancer risk factor examined (Additional file Table 1).

Next, a weighted average of study-specific estimates using inverse variance weights was derived for each risk factor [43] to increase the accuracy of outcomes [44, 45]. The potential for publication bias was evaluated by funnel plots and the methods described by Egger et al. [46] and Begg et al. [47]. Using a random-effects model, we analyzed the studies and considered heterogeneity and within-study variance [48]. We evaluated heterogeneity using Cochrane's Q-statistic [49] and the I^2 inconsistency statistical tests [50].

We considered the OR to be a good approximation of the RR for our analysis, which is reasonable when the outcome is rare [51]. We used the OR and the logOR and calculated standard errors (SEs) as data points for the meta-analysis. All statistical manipulations were conducted using the meta-analysis package for R (metaphor Version 2, MA, USA).

Steps 3a-3c: Analytic Hierarchy Process (AHP)

The third step in creating our index was to use the results of our meta-analysis as inputs for the AHP analysis and to generate weights for each risk factor. AHP is one of the most widely used Multi-Criterion Decision Making (MCDM) methods [52] and has been increasingly implemented in health care, including cancer research [53–57]. AHP can quantitatively prioritize risk factors

by producing weights for each factor, making it an ideal method to apply in this study. For each included modifiable risk factor, we used the OR derived from our meta-analysis as input variables in the AHP. Using the values from meta-analysis and the assessment matrix, we created the pair-wise comparison matrix (i.e., a matrix to compare risk factors in pairs to evaluate their relative importance). We created an assessment matrix with numbers that pair with different importance levels. For example, 1, 3, 5, 7, and 9 pair with equal, weak, obvious, intense, and extreme importance, while 2, 4, 6, and 8 pair with intermediate importance, respectively [58] (Additional file Table 1).

The relative importance of smoking versus all other included risk factors was assigned considering the assessment matrix. This step was then repeated for all other remaining risk factors. Next, an n by n matrix was created where n represented the number of modifiable risk factors. Next, we solved the linear system, where A is the coefficient matrix using Eq. 1:

$$AX = \lambda X \text{ or } (A - \lambda I_n)X = 0 \tag{1}$$

where A is the comparison matrix of order n , and λ is one of its eigenvalues. X represents the eigenvector of A associated with λ , and $A - \lambda I_n$ represents the matrix coefficient. We used MATLAB (MathWorks, Massachusetts, USA) to calculate the eigenvalues and eigenvectors of the matrix [59]. Then we used the derived eigenvector to specify the weights of each risk factor

$$LCRI = 0.461A_1 + 0.185A_2 + 0.132A_3 + 0.078A_4 + 0.076A_5 + 0.038A_6 + 0.030A_7 \tag{4}$$

where the eigenvector represented the index coefficient. Next, we estimated the contribution of each risk factor to lung cancer. We then calculated the z score and considered the z score as the corresponding value in the index. Finally, z -scores were converted to percentiles for mapping purposes.

We used the Consistency Ratio (CR) to verify the reliability of our results. To do this, we first calculated the Consistency Index (CI_1) using the following equation:

$$CI_1 = (\lambda_{max} - n) / (n - 1) \tag{2}$$

where λ_{max} was the maximum eigenvalue and n represented the order of the matrix. Accordingly, the CR was calculated by dividing the CI_1 by the index for the corresponding Random Index (RI) using the following equation:

$$CR_1 = CI_1 / RI \tag{3}$$

Table 1 Overall effect size and final weights for modifiable risk factors included in the Lung Cancer Risk Index (LCRI)

No	Risk factor	Overall effect size from meta-analysis	Weights used in LCRI (%)
1	Active smoking	8.63	46.1
2	Indoor air pollution	1.76	18.5
3	Occupational exposure	1.60	13.2
4	Alcohol consumption	1.45	7.8
5	Secondhand smoke exposure	1.43	7.6
6	Outdoor air pollution	1.25	3.8
7	Radon exposure	1.24	3.0
Total			100

Saaty [60] has presented the values for RI considering the matrix size. Also, Saaty [60] suggested that the CR needs to be less than 0.1 to produce consistent results.

Results

As shown in Table 1, the process that we used to create the LCRI yielded the highest weight for active smoking (46.1%) and the lowest weight for radon exposure (3.0%). The CR of the AHP analysis for the present study was 0.07, well below the 0.1 cut point that demonstrates consistency of the analysis.

We used the weights in Table 1 to produce the LCRI:

where A_1 to A_7 represent each included modifiable risk factor, as listed in Table 1. It should be noted that A_1 to A_7 can be values of 0 or 1, where 0 indicates the corresponding risk factor was not in effect and 1 indicates the corresponding risk factor was in effect (i.e., 0=no exposure and 1=exposure / risk exists). We calculated the corresponding z score for each geographical area (e.g., if the emitted air pollution for a county is X tons/year, the corresponding value for A_6 would be the corresponding z score which is dependent on the average and variance of emitted air pollution for that specific county compared to all other counties in any state). Developed countries such as the US do not rely on major sources of household air pollution—kerosene, wood, or coal—to generate heat [61, 62], so A_2 is assigned a value of 0 for individuals living in these countries. The LCRI can take any value between 0 and 1: an LCRI value of 0 means no predicted lung cancer risk (A_1 to A_7 all equal 0), and an LCRI value of 1 represents the highest possible predicted risk of lung cancer.

Case study

We test the adaptability and utility of the LCRI in a case study performed using data for our home state of Illinois. In this case study, we constructed the $LCRI_{IL}$ – a version of the LCRI that reflects the available population-level data in our state. IL is comprised of 102 counties, some of which are urban and many of which are rural. Forty percent of the state's population resides in Cook County – home to the City of Chicago. Cook County is the second-most populous county in the nation, with more than 5.2 million racially and ethnically diverse residents [63].

Our first step in creating the $LCRI_{IL}$ was to collect all necessary risk factor data from publicly available data sources. For all counties, we extracted data for 2014–2019 for active smoking (percentage of adults who are current smokers), radon exposure (pCi/L), outdoor air pollution (concentration of fine particulate matter (PM_{2.5})), and alcohol consumption (percentage of adults reporting binge or heavy drinking in past 30 days) [64, 65]. There were no publicly available county-level data for secondhand smoke exposure or occupational exposures, so those risk factors were dropped from the $LCRI_{IL}$.

The second step in creating the $LCRI_{IL}$ was to generate weights for each available risk factor using the previously described methods (see Methods, Steps 3a–3c). The weights used in the $LCRI_{IL}$ were 0.70 for active smoking, 0.14 for alcohol consumption, 0.095 for outdoor air pollution, and 0.057 for radon exposure. The corresponding equation to derive the $LCRI_{IL}$ is:

$$LCRI_{IL} = 0.701B_1 + 0.147B_2 + 0.095B_3 + 0.057B_4 \quad (5)$$

where B_1 to B_4 represent active smoking, alcohol consumption, outdoor air pollution, and radon exposure, respectively. The CR of the AHP analysis for the case study was 0.04, which indicated the consistency of the analysis.

Figure 3 shows the prevalence of each individual risk factor that was included in the $LCRI_{IL}$, as well as lung cancer outcomes [66], by county across Illinois. There is substantial heterogeneity for each risk factor across the state. Among the top 28 counties that have the highest lung cancer incidence and / or mortality rates, eight are also among the top 20 $LCRI_{IL}$ counties. These eight counties are predominantly located in rural areas (as defined by the US census, [63]) of Southern and Southeastern Illinois, though one is an urban county located on the east side of the state. Notably, Cook County had the highest $LCRI_{IL}$ score but among the lowest lung cancer incidence and mortality rates.

Table 2 presents Pearson correlation coefficients between the $LCRI_{IL}$ z scores, active smoking, and lung cancer incidence and mortality rates. The correlation

coefficients between the $LCRI_{IL}$ and lung cancer incidence and mortality were 0.45 and 0.50, respectively, with both p -values < 0.05. The correlation coefficient between the $LCRI_{IL}$ and active smoking was high at 0.87, which was expected given that this risk factor had the highest assigned weight in the index.

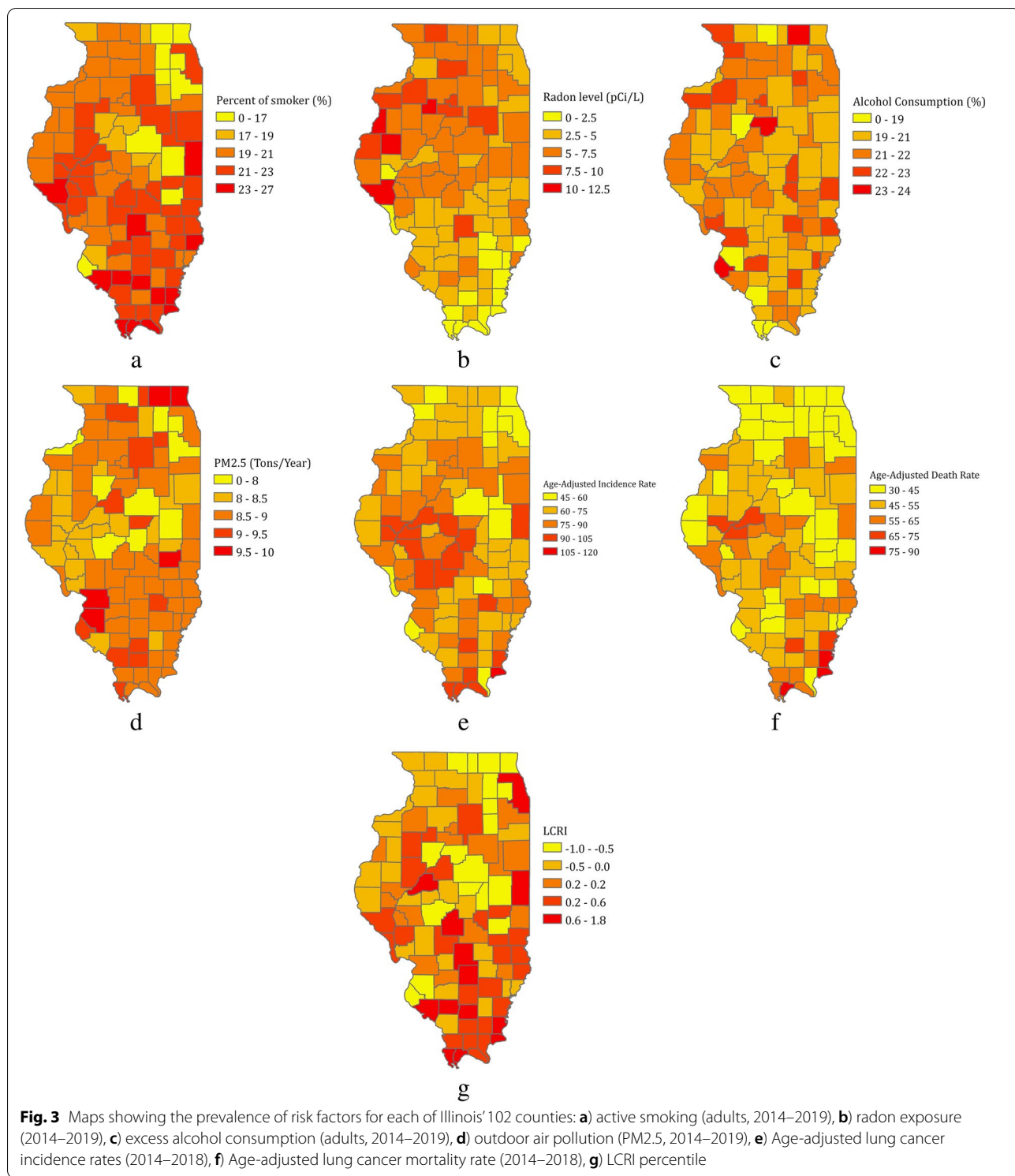
In sensitivity analyses, we examined the magnitude of the correlation coefficient for each component of the $LCRI_{IL}$ in relation to lung cancer incidence and mortality rates. The correlation coefficient was only statistically significant for active smoking, and the magnitude and significance were similar to that of the $LCRI_{IL}$ (Table 3). In an additional sensitivity analysis, alcohol consumption was dropped from the $LCRI_{IL}$ – since it is so highly correlated with active smoking – and the resulting index showed similar correlation with lung cancer incidence and mortality rates (0.496 and 0.545, respectively) as compared to the original index.

Discussion

We created a novel lung cancer risk index (LCRI) that integrates multiple modifiable risk factors into one measure. Active smoking is the predominant risk factor for lung cancer and is linked with 80–90% of lung cancer deaths [25]. As expected, smoking received the highest weight in both our original index (LCRI: 46.1%) and the one that we adapted for use in the state of Illinois ($LCRI_{IL}$: 70.1%). Conversely, radon exposure had the lowest weight in each index (LCRI: 3%, $LCRI_{IL}$: 5.7%).

Previous studies have largely focused on associations between individual risk factors and lung cancer outcomes [11, 13, 25, 29]. However, there are laboratory, animal, and human data showing that risk factors interact with each other to affect cancer outcomes [67–69]. For example, Wu et al. [67] reviewed and highlighted the evidence that cancer causation is multifactorial and suggested that researchers consider the contributions of individual factors and their joint effects on cancer burden. Li et al. showed that gene-smoking interactions play important roles in the etiology of lung cancer [68]. Our index represents an attempt to address these known interactions by using population-based data to capture the combined impact of multiple risk factors for lung cancer into one measure.

Hot spots identified by our index share similar distribution patterns of risk factors from the geospatial analysis. Interestingly, Cook County has the highest $LCRI_{IL}$ despite low adjusted lung cancer incidence and mortality rates. Although Cook County has moderate to high levels of alcohol consumption, Fine Particulate Matter 2.5, and air pollution, it also has a high ratio of primary care physicians to the population (1050:1, ranked



8th in IL), suggesting greater availability of healthcare resources. This may explain the discordance between Cook County's $LCRI_{IL}$ and lung cancer incidence and mortality rates. Counties with high $LCRI_{IL}$ and high lung cancer incidence or mortality rates are mostly in

the rural area of the state with fewer available healthcare resources [70]. This echoes findings from recent studies that cancer mortality rates associated with modifiable risks were higher in rural compared with urban populations [71, 72].

Table 2 Results of Pearson correlation test between LCRI_{IL}, percentage of active smokers, age-adjusted lung cancer incidence rate, and age-adjusted lung cancer mortality rate

	LCRI _{IL} z score	Active Smoker (%)	Age-Adjusted lung cancer Incidence Rate (2014–2018)	Age-Adjusted lung cancer mortality Rate (2014–2018)
LCRI _{IL} z score	1	.869 ^a	.449 ^a	.495 ^a
Active Smoker (%)		1	.498 ^a	.418 ^a
Age-Adjusted lung cancer Incidence Rate (2014–2018)			1	.812 ^a
Age-Adjusted lung cancer mortality Rate (2014–2018)				1

^a Correlation is significant at the 0.05 level (2-tailed)

Table 3 Sensitivity analysis of individual components of the LCRI_{IL} in relation to lung cancer outcomes in Illinois

		Age-adjusted IL lung cancer incidence rate (2014–2018)	Age-adjusted lung cancer mortality rate (2014–2018)
Active smoking	Pearson Correlation	.498 ^a	.418 ^a
	Sig. (2-tailed)	0.000	0.000
	N	102	102
Air pollution (PM2.5)	Pearson Correlation	0.066	0.055
	Sig. (2-tailed)	0.510	0.581
	N	102	102
Radon (pCi/L)	Pearson Correlation	-.124	-.126
	Sig. (2-tailed)	0.510	0.581
	N	102	102
Excess alcohol consumption (%)	Pearson Correlation	-.188	-0.063
	Sig. (2-tailed)	0.825	-.187
	N	102	102

^a Correlation is significant at the 0.05 level (2-tailed)

Cancer is a heterogeneous disease [73] with many risk factors at individual and social levels. Our model included the factors studied in the literature where the studies met the criteria for inclusion (e.g., being a modifiable risk factor, having an OR or RR, etc.); however, it is important to note that other non-modifiable factors such as age, gender, and race have been shown to also be strongly associated with lung cancer's incidence and mortality rates [74]. Nevertheless, the study offers a useful framework that health policymakers and researchers can use to identify and examine potential lung cancer risk factors for their geographical areas.

Our study has several strengths. First, to our knowledge, ours is the first study to use meta-analysis in combination with AHP to create a composite risk index for a specific cancer. Second, our model summarized complex and multi-dimensional factors to provide a tool for use by healthcare decision-makers. Our index includes several major and minor modifiable risk factors rather than

a single biomedical factor. Third, our study presents a new approach where researchers and policymakers can utilize databases (e.g., U.S. Centers for Disease Control & Prevention's Behavioral Risk Factor Surveillance System, U.S. Environmental Protection Agency's Office of Air Quality Planning and Standards, etc.) at multiple geographic levels to identify areas that may benefit from resource allocation and public health interventions. Additionally, a Meta-AHP approach could potentially be combined with machine learning and deep learning models [75, 76] to analyze risk factors and predict health outcomes more accurately.

There were several limitations to this study. First, the AHP approach only allows for the inclusion of risk estimates greater than 1. As a result, we could not include protective behaviors such as fruit and vegetable consumption in our index. Second, AHP relies directly and exclusively on the magnitude of a single risk estimate generated from the meta-analysis, which is likely an

underestimate because the model does not allow for variation in exposure prevalence by region. As an example, radon is widely considered to be the second leading cause of lung cancer, behind cigarette smoking [77]. However, as shown in Table 1, this risk factor received the lowest weight in the index because the risk estimate from the meta-analysis was only 1.24—the smallest magnitude of any factor examined. Third, we could not include secondhand smoke and occupational exposures in our tailored LCRI_{IL} index because county-level data in Illinois are not publicly available for these two factors. We also did not include non-modifiable risk factors such as age, gender, and race. Fourth, because alcohol consumption and tobacco smoking are strongly correlated, the confounding effect of smoking may impact the weight of alcohol consumption in the LCRI. However, when we removed alcohol consumption from LCRI_{IL} in a sensitivity test, the resulting index showed similar correlation to lung cancer outcomes. Future research is needed to examine the effect the strong correlation between smoking and alcohol has on the LCRI. Fifth, we imposed a single cut point for each risk factor in our models, while, in actuality, some risk factors may exhibit curvilinear or other types of relationships with cancer outcomes. Finally, the meta-analysis was limited to literature published in 1990 and beyond, and therefore did not capture earlier studies.

Conclusion

We generated a lung cancer risk index that incorporated several modifiable risk factors into one composite score. The index was driven heavily by active smoking, as expected. In addition, the index was modestly correlated with lung cancer outcomes in a case study conducted in Illinois, demonstrating its adaptability and potential utility in numerous geographic locations and potentially in many different fields. Future refinements to the index could include adding other modifiable risk factors, examining the impact of non-modifiable risk factors such as age, gender, and race / ethnicity in the LCRI, performing geographical cluster analysis, and incorporating other health behavior factors in AHP-based cancer risk factor models for lung cancer or other health outcomes.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12885-022-10370-4>.

Additional file 1.

Acknowledgements

Not applicable.

Authors' contributions

MCH, MV, MW, and AF conceptualized the study. All authors contributed to the study design. Material preparation, data collection and analysis were performed by AF and LG. The first draft of the manuscript was written by AF and LG and all authors commented on subsequent versions of the manuscript. All authors read and approved the final manuscript.

Funding

The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

Availability of data and materials

All data generated or analyzed during this study are included in this published article and its supplementary information files.

Declarations

Ethics approval and consent to participate

Not applicable. This is a secondary analysis using publicly available dataset.

Consent for publication

Not applicable.

Competing interests

The authors have no conflicts of interest to disclose.

Author details

¹College of Engineering and Engineering Technology, Northern Illinois University, DeKalb, IL, USA. ²School of Interdisciplinary Health Professions, Northern Illinois University, DeKalb, IL, USA. ³University of Illinois Cancer Center, Chicago, IL, USA. ⁴School of Health Studies, Northern Illinois University, DeKalb, IL, USA.

Received: 30 May 2022 Accepted: 25 November 2022

Published online: 06 December 2022

References

1. American Cancer Society; <https://www.cancer.org/>. [Cited 20 Apr 2021]. Available from
2. Malhotra J, et al. (2016) Risk factors for lung cancer worldwide. *Eur Respir J*. 2016;48(3):889–902.
3. Wang Q, Gümüş ZH, Colarossi C, Memeo L, Wang X, Kong CY, Boffetta P. Small Cell Lung Cancer: Epidemiology, Risk Factors, Genetic Susceptibility, Molecular Pathology, Screening and Early Detection. *Journal of thoracic oncology : official publication of the International Association for the Study of Lung Cancer*. 2022;S1556–0864(22)01851–2. Advance online publication.
4. Gudenkauf FJ, Thrift AP. Preventable causes of cancer in Texas by race/ethnicity: major modifiable risk factors in the population. *PLoS ONE*. 2022;17(10):e0274905.
5. Brenner DR, et al. Lung cancer risk in never-smokers: a population-based case-control study of epidemiologic risk factors. *BMC Cancer*. 2020;10(1):1–9.
6. Ridge CA, McErlean AM, Ginsberg MS. Epidemiology of lung cancer. in *Seminars in interventional radiology* Thieme Medical Publishers. 2013.
7. Couraud S, et al. Lung cancer in never smokers—a review. *Eur J Cancer*. 2012;48(9):1299–311.
8. Weiderpass E. Lifestyle and cancer risk. *J Prev Med Public Health*. 2010;43(6):459–71.
9. Dresler C. The changing epidemic of lung cancer and occupational and environmental risk factors. *Thorac Cardiovasc Surg*. 2013;23(2):113–22.
10. Samet JM, et al. Lung cancer in never smokers: clinical epidemiology and environmental risk factors. *Clin Cancer Res*. 2009;15(18):5626–45.
11. O'Keefe LM. Smoking as a risk factor for lung cancer in women and men: a systematic review and meta-analysis. *BMJ Open*. 2018;8(10):e021611.
12. Torre LA, Siegel RL, Jemal A. Lung cancer statistics. *lung cancer and personalized medicine*. *Adv Exp Med Biol*. 2016;893:1–19.
13. Taylor R, Najafi F, Dobson A. Meta-analysis of studies of passive smoking and lung cancer: effects of study type and continent. *Int J Epidemiol*. 2007;36(5):1048–59.

14. Gawełek E, Drozdowska B, Fuchs A. Radon as a risk factor of lung cancer. *Przegl Epidemiol*. 2017;71(1):90–8.
15. VENA, J.E. Air pollution as a risk factor in lung cancer. *Am J Epidemiol*. 1982;116(1):42–56.
16. Behera D, Balamugesh T. Indoor air pollution as a risk factor for lung cancer in women. *JAPI*. 2005;53:190–2.
17. Gustavsson P, et al. Occupational exposure and lung cancer risk: a population-based case-referent study in Sweden. *Am J Epidemiol*. 2000;152(1):32–40.
18. Brennan P, et al. A multicenter case–control study of diet and lung cancer among non-smokers. *Cancer Causes Control*. 2020;11(1):49–58.
19. Bandera EV, Freudenheim JL, Vena JE. Alcohol consumption and lung cancer: a review of the epidemiologic evidence. *Cancer Epidemiol Biomarkers Prev*. 2001;10(8):813–21.
20. Kanwal M, Ding XJ, Cao Y. Familial risk for lung cancer. *Oncol Lett*. 2017;13(2):535–42.
21. Wu AH, et al. Previous lung disease and risk of lung cancer among lifetime nonsmoking women in the United States. *Am J Epidemiol*. 1995;141(11):1023–32.
22. Celik I, et al. Arsenic in drinking water and lung cancer: a systematic review. *Environ Res*. 2008;108(1):48–55.
23. Świętkowska B. Modifiable risk factors for the prevention of lung cancer. *Rep Prac Oncol Radiother*. 2007;12(2):119–24.
24. Turner MC, et al. Radon and lung cancer in the American Cancer Society cohort. *Cancer Epidemiol Prev Biomark*. 2011;20(3):438–48.
25. Yun YH, et al. Relative and absolute risks of cigarette smoking on major histologic types of lung cancer in Korean men. *Cancer Epidemiol Prev Biomark*. 2005;14(9):2125–30.
26. Nikić D, Stanković AM. Air pollution as a risk factor for lung cancer. *Arch Oncol*. 2005;13(2):79–82.
27. Brenner DR, et al. Alcohol consumption and lung cancer risk: A pooled analysis from the International Lung Cancer Consortium and the SYNERGY study. *Cancer Epidemiol*. 2019;58:25–32.
28. Kreienbrock L, et al. Case-control study on lung cancer and residential radon in western Germany. *Am J Epidemiol*. 2001;153(1):42–52.
29. Pesch B, et al. Cigarette smoking and lung cancer—relative risk estimates for the major histological types from a pooled analysis of case–control studies. *Int J Cancer*. 2012;131(5):1210–9.
30. Lee T, Gany F. Cooking oil fumes and lung cancer: a review of the literature in the context of the US population. *J Immigr Minor Health*. 2013;15(3):646–52.
31. Stockwell HG, et al. Environmental tobacco smoke and lung cancer risk in nonsmoking women. *J Natl Cancer Inst*. 1992;84(18):1417–22.
32. Miller AB, et al. Fruits and vegetables and lung cancer: findings from the European prospective investigation into cancer and nutrition. *Int J Cancer*. 2004;108(2):269–76.
33. Menvielle G, et al. The role of smoking and diet in explaining educational inequalities in lung cancer incidence. *J Natl Cancer Inst*. 2009;101(5):321–30.
34. Spitz MR, et al. A risk model for prediction of lung cancer. *J Natl Cancer Inst*. 2007;99(9):715–26.
35. Xue R, Wang C, Liu M, Zhang D, Li K, Li N. A new method for soil health assessment based on Analytic Hierarchy Process and meta-analysis. *Sci Total Environ*. 2019;650:2771–7.
36. Liberati A, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *J Clin Epidemiol*. 2009;62(10):e1–34.
37. World Cancer Research Fund 2021; Lung Cancer. <https://www.wcrf.org/dietandcancer/lung-cancer/>. [Cited 1 Jun 2021]. Available from
38. Pontius FW. Appendix G: Listing of Drinking Water Federal Register Notices. *Drinking Water Regulation and Health*. 2003:53–969.
39. Federal Register. National Primary Drinking Water Regulations; Arsenic and Clarifications to Compliance and New Source Contaminants Monitoring]. <https://www.federalregister.gov/documents/2001/01/22/01-1668/national-primary-drinking-water-regulations-arsenic-and-clarifications-to-compliance-and-new-source>. [Cited 1 Jun 2021]. Available from
40. Boffetta, and C. Borron., Low-level exposure to arsenic in drinking water and risk of lung and bladder cancer: a systematic review and dose–response meta-analysis. *Dose-Response*. 2019;17(3):1559325819863634.
41. JBI's Critical Appraisal Tools. 2021. [Cited 1 Mar 2021]. Available from <https://jbi.global/critical-appraisal-tools>.
42. Page MJ, et al. (2021) PRISMA 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews. *BMJ*. 2021;372:160.
43. Kleinbaum D, Kupper L, Morgenstern H. Principles and quantitative methods. *Epidemiologic research*. New York: Van Nostrand Reinhold Company; 1982. p. 427–50.
44. Borenstein M. Introduction to meta-analysis. John Wiley & Sons; 2021.
45. Hunter JE, Schmidt FL. Methods of meta-analysis: Correcting error and bias in research findings. Sage; 2004.
46. Egger M, et al. Bias in meta-analysis detected by a simple, graphical test. *BMJ*. 1997;315(7109):629–34.
47. Begg CB, Mazumdar M. Operating characteristics of a rank correlation test for publication bias. *Biometrics*. 1994;50(4):888–1101.
48. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials*. 1986;7(3):177–88.
49. Cochran WG. The combination of estimates from different experiments. *Biometrics*. 1954;10(1):101–29.
50. Higgins JP, et al. Measuring inconsistency in meta-analyses. *BMJ*. 2003;327(7414):557–60.
51. Matakidou A, Eisen T, Houlston R. Systematic review of the relationship between family history and lung cancer risk. *Br J Cancer*. 2005;93(7):825–33.
52. Mu E, Pereyra-Rojas M. Understanding the analytic hierarchy process, in Practical decision making Springer. 2017;7–22.
53. Schmidt K, et al. Applying the Analytic Hierarchy Process in healthcare research: a systematic literature review and evaluation of reporting. *BMC Med Inform Decis Mak*. 2015;15(1):1–27.
54. Hummel JM, Bridges J, Ilzerman, M. J. Group decision making with the analytic hierarchy process in benefit-risk assessment: a tutorial. *The Patient*. 2014;7(2):129–40.
55. Sakti CY, Sungkono KR, Sarno R. International Seminar on Application for Technology of Information and Communication (iSemantic). Determination of Hospital Rank by Using Analytic Hierarchy Process (AHP) and Multi Objective Optimization on the Basis of Ratio Analysis (MOORA). 2019:178–183.
56. Schmidt K, Aumann I, Hollander I, Damm K, Schulenburg J-MG. Applying the Analytic Hierarchy Process in healthcare research: a systematic literature review and evaluation of reporting. *BMC Med Inform Decis Mak*. 2015;1:234. <https://doi.org/10.1186/s12911-015-0234-7>.
57. Pauer F, Schmidt K, Babac A, Damm K, Frank M, von der Schulenburg JM. Comparison of different approaches applied in Analytic Hierarchy Process - an example of information needs of patients with rare diseases. *BMC Med Inform Decis Mak*. 2016;16(1):117. <https://doi.org/10.1186/s12911-016-0346-8>.
58. Saaty TL. Decision making with the analytic hierarchy process. *Int J Serv Sci*. 2008;1(1):83–98.
59. Moler CB. Numerical computing with MATLAB. SIAM; 2004.
60. Saaty, T.L. (1989) Group decision making and the AHP, in The analytic hierarchy process. Springer. pp. 59–67.
61. Bruce N, Perez-Padilla R, Albalak R. Indoor air pollution in developing countries: a major environmental and public health challenge. *Bull World Health Organ*. 2000;78(9):1078–92.
62. WEO-2017 Special Report: Energy Access Outlook, International Energy Agency, 2017 <https://webstore.iea.org/weo-2017-special-report-energy-access-outlook>. Available from
63. 2020 United States census
64. Illinois Department Of Public Health Division Of Behavioral Risk Factor Surveillance System. <http://www.idph.state.il.us/brfss/>. [Cited Aug 2020]. Available from
65. Illinois Department of Public Health. <http://www.idph.illinois.gov>. Available from
66. United States Cancer Statistics 2021. Centers for Disease Control and Prevention]. <https://gis.cdc.gov/cancer/uscs/#/StateCounty/>. [Cited 1 Aug 2021]. Available from
67. Wu S, Zhu W, Thompson P, Hannun YA. Evaluating intrinsic and non-intrinsic cancer risk factors. *Nat Commun*. 2018;9(1):3490. <https://doi.org/10.1038/s41467-018-05467-z>.
68. Li Y, Xiao X, Han Y, Gorlova O, Qian D, Leigh N, Johansen JS, Barnett M, Chen C, Goodman G, Cox A, Taylor F, Woll P, Wichmann HE, Manz J, Muley T, Risch A, Rosenberger A, Arnold SM, Haura EB, Amos CI. Genome-wide interaction study of smoking behavior and non-small cell lung cancer risk

- in Caucasian population. *Carcinogenesis*. 2018;39(3):336–46. <https://doi.org/10.1093/carcin/bgx113>.
69. Biesalski HK, de MesquitaBuenoChesson BA, Chytil F, Grimble R, Hermus RJ, Köhrle J, Lotan R, Norpoth K, Pastorino U, Thurnham D. European consensus statement on lung cancer: risk factors and prevention lung cancer panel. *CA Cancer J Clin*. 1998;48(3):167–166.
 70. Health Resources and Administration (HRSA). Rural Access to Health Care Services Request for Information, available from <https://www.hrsa.gov/rural-health/rfi-rural-health-care-access>
 71. Fogleman AJ, Mueller GS, Jenkins WD. Does where you live play an important role in cancer incidence in the U.S.? *Am J Cancer Res*. 2015;5(7):2314–9.
 72. Moss JL, Pinto CN, Srinivasan S, Cronin KA, Croyle RT. Persistent poverty and cancer mortality rates: an analysis of county-level poverty designations. *Cancer Epidemiol Biomarkers Prev*. 2020;29(10):1949–54. <https://doi.org/10.1158/1055-9965.EPI-20-0007>.
 73. Hiatt RA, Breen N. The social determinants of cancer: a challenge for transdisciplinary science. *Am J Prev Med*. 2008;35(2 Suppl):S141–50. <https://doi.org/10.1016/j.amepre.2008.05.006>.
 74. Stram DO, Park SL, Haiman CA, Murphy SE, Patel Y, Hecht SS, Le Marchand L. Racial/ethnic differences in lung cancer incidence in the multiethnic cohort study: an update. *J Natl Cancer Inst*. 2019;111(8):811–9.
 75. Afshar P, Mohammadi A, Tyrrell N, Cheung P, Sigiuk A, Plataniotis KN, Nguyen ET, Oikonomou A. [Formula: see text]: deep learning-based radiomics for the time-to-event outcome prediction in lung cancer. *Sci Rep*. 2020;10(1):12366. <https://doi.org/10.1038/s41598-020-69106-8>.
 76. Lima E, Gorski E, Loures EFR, Portela SEA, Deschamps F, & 9th IFAC Conference on Manufacturing Modelling, Management and Control, MIM 2019. (2019). Applying machine learning to AHP multicriteria decision making method to assets prioritization in the context of industrial maintenance 4.0. *Ifac-papersonline*, 52, 13, 2152–2157
 77. Health risk of Radon. <https://www.epa.gov/radon/health-risk-radon>. Available from

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

