Published in final edited form as:
Int J Cancer. 2019 June 15; 144(12): 2918–2927. doi:10.1002/ijc.32034.

Lung cancer risk by geologic coal deposits: A case–control study of female never-smokers from Xuanwei and Fuyuan, China

Jason Y.Y. Wong[1], **George S. Downward**[2], **Wei Hu**[1], **Lützen Portengen**[2], **Wei Jie Seow**[3,4], **Debra T. Silverman**[1], **Bryan A. Bassig**[1], **Jinming Zhang**[1], **Jun Xu**[5], **Bu-Tian Ji**[1], **Jihua Li**[6], **Jun He**[6], **Kaiyun Yang**[7], **Linwei Tian**[5], **Min Shen**[8], **Yunchao Huang**[7,*], **Roel Vermeulen**[2,*], **Nathaniel Rothman**[1,*], **Qing Lan**[1,*]

[1]Occupational and Environmental Epidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Rockville, MD, USA

[2]Institute for Risk Assessment Sciences, Division of Environmental Epidemiology, Utrecht University, Utrecht, The Netherlands

[3]Saw Swee Hock School of Public Health, National University of Singapore, Singapore, Singapore

[4]Department of Medicine, Yong Loo Lin School of Medicine, National University of Singapore and National University Health System, Singapore, Singapore

[5]School of Public Health, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Hong Kong, Special Administrative Region of China

[6]Qujing Center for Diseases Control and Prevention, Qujing, Yunnan, China

[7]Kunming Tumor Hospital, Kunming, Yunnan, China

[8]Caromont Inpatient Physicians, Gastonia, NC, USA

Abstract

Coal types vary around the world because of geochemical differences in their source deposits; however, the influence of coal emissions from different deposits on human health remains unexplored. To address this issue, we conducted the first study of the relationship between coal use from various deposits and lung cancer risk in Xuanwei and Fuyuan, counties in China where lung cancer rates are among the highest in the world among female never-smokers due to use of bituminous ("smoky") coal for heating and cooking. We conducted a population-based case–control study of 1031 lung cancer cases and 493 controls among never-smoking women in Xuanwei and Fuyuan. Logistic regression models were used to estimate associations between coal use from various deposits across the lifecourse and lung cancer risk. There was substantial heterogeneity in risks by coal deposit ($p = 7.8E-05$). Compared to non-smoky coal users, risks by smoky coal deposit ranged from $\mathrm{OR} = 7.49$ (95% CI: $3.43-16.38$) to

---

Correspondence to: Jason Y.Y. Wong, National Cancer Institute, 9609 Medical Center Drive, Rockville, MD 20850, USA, Tel.: 1-240-276-5149, Fax: 1-240-276-5149, jason.wong@nih.gov.

[*] Y.H., R.V., N.R., and Q.L. co-supervised the study.

Additional Supporting Information may be found in the online version of this article.

$\mathrm{OR} = 33.40\ (95\%\ \mathrm{CI}{:}\ 13.07 - 85.34)$. Further, women born into homes that used smoky coal and subsequently changed to non-smoky coal had a higher risk $(\mathrm{OR} = 10.83\ (95\%\ \mathrm{CI}{:}\ 4.61 - 25.46))$ than women born into homes that used non-smoky coal and changed to smoky coal $(\mathrm{OR} = 4.74\ (95\%\ \mathrm{CI}{:}\ 2.03 - 11.04,\ p_{\mathrm{difference}} = 0.04))$. Our study demonstrates that various sources of coal have considerably different impact on lung cancer in this population and suggests that early-life exposure to carcinogenic emissions may exert substantial influence on health risks later in life. These factors should be considered when evaluating the health risks posed by exposure to coal combustion emissions.

Keywords

lung cancer; geologic coal deposit; bituminous “smoky” coal; geographic variation; indoor air pollution

Introduction

Nearly half of the world’s population is exposed to household air pollution from the combustion of solid fuels such as coal for heating and cooking.[1] Indoor coal combustion emissions have been designated as Group 1 carcinogens by a working group of the International Agency for Research on Cancer (IARC).[2–4] The carcinogenic effect of coal combustion emissions exhibits global geographic variation, with some of the highest rates of lung cancer found in the rural county of Xuanwei located in southwestern China.[5] Residents of neighboring Fuyuan county have comparably high lung cancer rates.[6] Previous studies suggest that this health burden is largely attributed to household combustion of a particularly toxic form of bituminous (“smoky”) coal that is locally mined.[7–10] Women from Xuanwei and Fuyuan are predominantly nonsmokers but have exceptionally high exposure to residential smoky coal emissions because of their traditional agrarian lifestyle in which they spend considerable amounts of time indoors preparing food.[11,12] Their exposure is further exacerbated by cooking with unvented fire pits, which allows smoky coal emissions to permeate the home.

In a previously conducted population-based case–control study, we found variation in lung cancer risk by broad administrative region in Xuanwei,[11,13–15] with residents of Laibin township having among the highest risks. However, variation in lung cancer risk by specific geologic source deposit of household coal was not assessed at the time due to lack of information on their exact location.

Previous studies have suggested that the health consequences of coal use differ according to the geochemical subtype.[16,17] Subtypes of coal from different regions may have undergone unique mineralization conditions that may have led to the enrichment of potentially toxic trace elements.[16–18] Smoky coal found across Xuanwei may possess unique geochemical properties and origins compared to coal from other regions of the world, which might account for their high carcinogenicity.[18] The distinctiveness of geologic deposits across Xuanwei may reflect differences in coalification or depositional environment.[15] However, no study to our knowledge has investigated whether lung cancer risks differ by specific geologic source deposits of household coal. Additionally, the influence of different types of coal used in the home during childhood *versus* adulthood on lung cancer risk has yet to be investigated.

To study the etiology of lung cancer in this population and ultimately identify the characteristics of coal that account for the extraordinary risk, we conducted a population-based case–control study of lung cancer among female never-smokers. We investigated the influence of using coal from different geologic source deposits across the lifecourse on lung cancer risk.

Methods

Study design, case ascertainment, and control selection

We conducted a case–control study of lung cancer risk among female never-smokers from Xuanwei and Fuyuan, China between 2006 and 2013. Given the rural and semi-mountainous nature of the area, there is no cancer registry in this region. Therefore, we enrolled cases that were admitted to six hospitals that diagnosed the vast majority of reported lung cancer cases in Xuanwei and Fuyuan. Eligible incident lung cancer cases (International Classification of Disease 9th revision (ICD-9) code: 162) were newly-diagnosed female never-smokers, aged 18–79 years old at the time of diagnosis, currently living in Xuanwei or Fuyuan, had lived in these counties for at least 1 year, and who had no history of previous cancer diagnoses. Cases were initially identified based on clinical presentation and a chest X-ray and CT scan with a presumptive diagnosis of lung cancer. Subjects then provided multiple sputum samples for cytologic analysis and testing for tuberculosis (TB). A subgroup of cases had more extensive diagnostic testing. For cases without a confirmed diagnosis of lung cancer made by histology or cytology, clinical evidence, radiologic findings (i.e., chest X-ray, CT scan), and lack of evidence for TB, pneumonia, or other conditions were used as the basis for the initial diagnosis of lung cancer.

A radiologist, cytologist, and pathologist at Peking Union Hospital in Beijing reviewed medical records, chest X-rays, CT scans, cytology slides, and tumor tissue slides. For cases with a diagnosis based on histological analysis of tumor tissue, 81% were adenocarcinoma. Cases that were not confirmed based on diagnostic biological samples were followed-up for vital status and cause of death for at least 3 years after enrollment into the study and classified as “high probability” lung cancer cases if they did not have another diagnosis explaining their initial presentation or cause of death. A total of 1060 eligible confirmed and “high probability” lung cancer cases were considered. The participation rate for cases was 84.4%. We excluded 29 cases who were predicted to use coal deposits 11, 15, 21, 23, 25, 26, 29, 30, and 31 (Fig. 1, Supporting Information Table 1) because of sparse numbers and the inability to be combined with adjacent deposits due to geochemical distinctiveness, which left 1031 cases for the analyses. Among these cases, 689 (66.8%) were confirmed and 342 (33.2%) were “high probability” cases. Cases were interviewed using a detailed questionnaire to collect information on residential history, fuel use, and established or suspected risk factors for lung cancer. Additionally, cases were asked to provide sputum samples, a tumor sample if available, and additional biological samples including peripheral blood.

A total of 498 population-based controls who were never-smokers, did not have a previous lung cancer diagnosis, were currently living in Xuanwei or Fuyuan, had lived in these counties for at least 1 year, and did not have a history of cancer were enrolled in our study. The sample size was selected based on the very strong associations between smoky coal use and lung cancer risk in this region that were found in our previous studies.[7,13] The participation rate for population-based controls was 89.0%. Control sampling was based on four geographic levels from largest to smallest, as follows: (1) commune, (2) administrative villages ("dadui"), (3) natural villages/settlements, and (4) individuals. The population-based controls from each commune were frequency-matched to cases by age based on the population density of adult women in each commune. First, from the population density of each commune, we randomly selected administrative villages, followed by randomly selecting up to three natural villages/settlements. Subsequently, the field team of interviewers visited each natural village, selected three potential participants fitting the eligibility criteria, and randomly recruited one of them. Among the eligible population-based controls, we excluded 5 women who were predicted to use coal deposits 11, 15, 23, 25, 26, 31, 21, 29, and 30, which left 493 population-based controls for the analyses.

The study protocol was approved by the institutional review boards of the National Cancer Institute (NCI) and China National Environmental Monitoring Center. All participants provided written informed consent.

Coal deposits

Information on coal deposit location and geological subtype was provided by local geologists and subsequently geocoded against the location of participants' villages using ArcGIS. The geographic location of each coal deposit is depicted in Figure 1. The coal deposits were classified as smoky coal (i.e., coking coal, 1/3 coking coal, meager lean coal, gas fat coal) and smokeless coal based on the degree of coalification and caking property characteristics as defined by the State Standard of China Coal Classification.[19] Coal deposit use was imputed for each year of the participant's life as described in the Appendix.

Analyses

Combined coal deposit exposure groups/categories were created based on geographic proximity, coal subcategorization, and the number of women who used coal from each deposit (Fig. 1, Supporting Information Table 1): A) Deposits 1, 2, 4, 7, 8 (coking coal from northern Xuanwei); B) Deposit 9 (Laibin coking coal); C) Deposit 10 (coking coal from southern Xuanwei); D) Deposits 12, 13, 14, 38 (gas fat); E) Deposits 16, 17, 19 (coking coal from Fuyuan); F) Deposit 24 (Laibin meager lean); G) Deposit 20, 27 (ever smoky coal users who lived in smokeless coal producing regions); H) unknown and other/outside of Xuanwei and Fuyuan; I) wood and dung users from smoky coal deposit areas; and J) never smoky coal users from smokeless coal regions who used wood, dung, or smokeless coal deposits (reference). For separate analyses, we also combined all smoky coal deposits into a single group/category.

Lifetime tonnage of smoky coal use was dichotomized based on the median levels of the population-based controls (low: <125, high: ≥125). Combination variables were created between dichotomous tonnage and each smoky coal deposit group.

Differences in continuous and categorical variables between cases and controls were assessed using Wilcoxon rank sum tests and $\chi 2$ or Fisher's Exact tests, respectively. Unconditional multivariable logistic regression models were used to estimate the odds ratios (OR) and 95% confidence intervals (CI) of incident lung cancer in relation to coal deposit group, adjusted for continuous age (years) and pre-marriage food sufficiency (not enough, enough, more than enough). Given the economic instability in China from 1953 to 1976 during the initial Five-Year Plans, Great Leap Forward, and Cultural Revolution, pre-marriage food sufficiency was deemed *a priori* the most appropriate proxy for SES.

We analyzed the coal deposit group most frequently used across the lifecourse and at the participant's birth home as the main exposures in separate models. The participants' county of residence (a frequency-matching factor) was inherently reflected by the predicted coal deposits. Variation in risk estimates by coal deposit groups/categories was assessed using $\chi 2$ tests.[20] Contrasts of deposit groups/categories were performed by changing the reference group of the indicator variable in the logistic regression models.

We conducted additional analyses of lung cancer risk in relation to: A) dichotomous lifetime tonnage; B) a combination variable between dichotomous lifetime tonnage and coal deposit group most frequently used across the lifecourse and at the birth home; C) the most frequently used coal deposit group with a 5- and 10-year exposure lag; and D) permanent change from smoky coal used in the birth home to non-smoky coal fuels, or from non-smoky coal fuels used in the birth home to smoky coal. Further sensitivity analyses were conducted by: A) restricting to confirmed lung cancer cases and B) excluding those with a family history of lung cancer.

We considered and evaluated other potential confounders/covariates for inclusion in the analyses. History of respiratory disease was not included because it was considered a mediator on the causal pathway between coal use and lung cancer development.[21,22] Family history of lung cancer was not included because it was highly correlated with coal deposit use from shared family environment. Secondhand smoke exposure was not included due to lack of statistical variability. Cumulative time spent indoors, ever used a stove with a chimney, menopausal status, and age at which the women began cooking were not included in the final models because they were either not associated with lung cancer or their inclusion did not appreciably change the findings.

All analyses were conducted using SAS v9.3 (SAS Institute Inc., Cary, NC, USA). p-Values <0.05 were considered statistically significant.

Results

Coal deposit use and study population characteristics

The distribution of case and control characteristics is shown in Table 1, while the distribution of coal deposits is shown in Supporting Information Table 1. The average age of both cases and controls was 54.7 years (Table 1). Most cases were from Xuanwei (61.4%) and had ever used smoky coal (94.4%). Cases used a substantially greater amount of smoky coal compared to controls (178 *vs.* 148 tons). Additionally, cases reported a higher proportion of having more than enough food before marriage (11.3% *vs.* 6.1%) compared to controls. Cases also had a greater proportion of first-degree relatives with lung cancer (19.9% *vs.* 6.5%) and history of respiratory disease (10.1% *vs.* 4.1%) compared to controls.

Associations between coal deposit use and lung cancer risk

Overall, the use of smoky coal deposits across the lifecourse was associated with substantially increased risks of developing lung cancer compared to never use (Table 2). Further, there was evidence that the risk estimates differed by deposit groups from which the coal was sourced ($p = 7.8\mathrm{E}-05$). Women who most frequently used smoky coal from Deposits 9 and 24 in Laibin township of Xuanwei county had the highest risks [$\mathrm{OR} = 33.40$ $(95\%\ \mathrm{CI}: 13.07-85.34,\ p = 2.3\mathrm{E}-13)$ and $\mathrm{OR} = 20.86$ $(95\%\ \mathrm{CI}: 7.16-60.78,\ p = 2.6\mathrm{E}-08)$, respectively]. The highest risks in Fuyuan county were found for Deposits 12, 13, 14, and 38 combined and Deposits 16, 17, and 19 combined [$\mathrm{OR} = 16.32$ $(95\%\ \mathrm{CI}: 6.26-42.54,\ p = 1.1\mathrm{E}-08)$ and $\mathrm{OR} = 14.27$ $(95\%\ \mathrm{CI}: 7.07-28.78,\ p = 1.2\mathrm{E}-13)$, respectively]. In contrast, women who used smoky coal from Deposit 10 in southern Xuanwei had a considerably lower risk $(\mathrm{OR} = 7.49$ $(95\%\ \mathrm{CI}: 3.43-16.38,\ p = 4.6\mathrm{E}-07)$ compared to the elevated risk associated with use of smoky coal from Laibin township ($p_{\mathrm{difference}} = 3.6\mathrm{E}-04$).

We found similar overall trends for lung cancer risk for coal deposit group used at the birth home (Table 2), with similar evidence for variation in risks by the deposits in which the coal was sourced ($\chi 2\ p = 1.4\mathrm{E}-04$). In their birth home, women who used smoky coal deposits from Laibin township of Xuanwei (Deposit 9: $(\mathrm{OR} = 24.15$ $(95\%\ \mathrm{CI}: 10.45-55.81,\ p = 9.5\mathrm{E}-14)$ and Deposit 24: $(\mathrm{OR} = 24.08$ $(95\%\ \mathrm{CI}: 7.08-81.89,\ p = 3.5\mathrm{E}-07))$ and Fuyuan (Deposits 12, 13, 14, 38: $(\mathrm{OR} = 34.08$ $(95\%\ \mathrm{CI}: 11.14-104.26,\ p = 6.2\mathrm{E}-10)$ and Deposits 16, 17, 19: $(\mathrm{OR} = 17.81$ $(95\%\ \mathrm{CI}: 8.74-36.29,\ p = 2.2\mathrm{E}-15)]$ were found to have the highest risks. Use of smoky coal from Deposit 10 in southern Xuanwei had a lower risk $(\mathrm{OR} = 8.56$ $(95\%\ \mathrm{CI}: 3.82-19.17,\ p = 1.8\mathrm{E}-07)$ compared to smoky coal from Deposit 9 in Laibin ($p_{\mathrm{difference}} = 5.5\mathrm{E}-04$). Similar patterns were found when analyzing a 5- and 10-year exposure lag (Supporting Information Table 2), when analyzing only confirmed cases of lung cancer (data not shown), and when excluding those with a family history of lung cancer (data not shown).

Relationship between cumulative tons of smoky coal used, coal deposit, and lung cancer risk

Women who used ≥125 cumulative tons of smoky coal across their lifetime had a 1.42 $(95\%\ \mathrm{CI}:\ 1.07-1.89,\ p=1.6\mathrm{E}-02)$ times increased risk of lung cancer compared to those who used <125 tons. Further, the risks were consistently greater for women having used higher *vs.* lower amounts of smoky coal tons for each deposit group (Table 3). These trends were similar in analyses of coal deposits used in the birth home (Table 3).

Change in coal/fuel exposure in birth home vs. later in life and lung cancer risk

Women born into homes where smoky coal was used as the predominant fuel type who permanently changed to other fuel types (i.e., smokeless coal, wood, dung, gas or electricity) had a lung cancer risk of $\mathrm{OR}=10.83\ (95\%\ \mathrm{CI}{:}4.61-25.46,\ p\ =\ 4.6\mathrm{E}-08)$ (Table 4). In contrast, women born into homes that used smokeless coal, wood, dung, gas or electricity as the predominant fuel type who permanently changed to smoky coal had a relatively lower lung cancer risk $(\mathrm{OR}=4.74\ (95\%\ \mathrm{CI}:\ 2.03-11.04,\ p\ =\ 3.2\mathrm{E}-04)\ (p_{\mathrm{difference}}=0.04)$ (Table 4). Cumulative smoky coal tonnage exposure for cases in each group were comparable (median = 128.0 tons for cases who changed from smoky coal at birth to other fuels *vs.* median = 143.5 tons for cases who changed from other fuels at birth to smoky coal).

Discussion

To our knowledge, this is the first study to investigate the variation in lung cancer risk by geologic coal deposit. We found that the use of smoky coal deposits across the lifecourse in Xuanwei and Fuyuan, China was associated with increased risks of lung cancer compared to never use of smoky coal, with evidence of variation by the source deposit that supplied coal to each household. The use of smoky coal from two deposits in Laibin township of Xuanwei were among the most hazardous. Moreover, we found that certain deposits from neighboring Fuyuan county were also associated with elevated risks comparable to those found in high-risk Laibin. Additionally, there was evidence for exposure-response trends between lifetime cumulative smoky coal use and lung cancer risk.

In contrast to other studies, we were able to evaluate household use of coal from various deposits during childhood. We found that women who were exposed to smoky coal at birth and then lived in homes that burned other fuels such as smokeless coal, wood, or dung had higher risks compared to women who were born into homes that used non-smoky coal sources of fuel who later lived in homes that used smoky coal. These findings add further etiologic insight to the literature on the pathogenesis of lung cancer in humans,[23] highlighting the importance of early life exposure to carcinogens and cancer risk in adulthood.[23,24]

Coal combustion emissions are complex mixtures of multiple components that could play a role in lung carcinogenesis. The hazardous components of these emissions include polycyclic aromatic hydrocarbons (PAHs) and related species including alkylated PAHs, fine and ultrafine particulate matter ($PM_{2.5}$, PM_1), carbon monoxide (CO), silica/quartz, sulfur oxides (SOx), and nitrogen oxides (NOx).[9,11,14,15,18,25–28] The precise constituent(s) of the smoky coal deposits that drive the strikingly high lung cancer risks and geographic variation in Xuanwei and Fuyuan are under investigation. Understanding the carcinogenic mechanism of smoky coal emissions is a key goal of the current study and we have conducted a detailed exposure study that will be used to model exposure to a wide range of potential carcinogens and estimate risk.[11,13–15,29]

Carcinogenic components of coal emissions can induce genomic damage in progenitor cells in early life during periods of rapid development, which can then propagate in subsequent generations of daughter cells into adulthood. Should multiple genomic perturbations occur in proto-oncogenes or tumor suppressors as people age, cancer development may be initiated and promoted.[23,30] Aside from genomic damage, components of coal emissions may cause other adverse biological effects including immune/inflammatory response,[31] adduct formation with proteins which can impair their molecular function,[32] and accelerated DNA methylation aging.[33]

The adverse health effects of coal combustion may be enhanced by early life exposure compared to exposures occurring later in life. Traffic-related air pollution contains certain common toxic constituents as coal emissions. A longitudinal twin study that measured telomere length in prenatal placenta and buccal cells in young adulthood found inverse associations between traffic-related air pollution and telomere length.[34] Interestingly, the authors found that residential traffic exposure at the birth address was significantly and inversely associated with telomere length ranking between birth and adulthood; however, residential traffic exposure in adulthood was not associated with telomere length.[34] Further, the authors found associations between residential traffic exposure at birth and accelerated telomere shortening in the first two decades of life.[34] There have been several longitudinal studies of the influence of traffic-related air pollution on lung function in early development *versus* later in life.[35,36] For example, a study of 2307 children from the Oslo Birth Cohort found that increased early life and lifetime exposure to PM_{10} and NO_2 were associated with reduced expiratory flow measurements, with exposure in the first year of life having slightly stronger effects.[37] Additionally, there have been efforts to characterize the influence of traffic-related air pollution across the lifecourse on cancer development. For instance, the Western New York Exposures and Breast Cancer (WEB) study assessed the associations between pre- and postmenopausal breast cancer risk and exposure to traffic emissions at menarche, first birth, and 10 and 20 years before interview using a case–control study design. The investigators found a significant exposure-response trend for exposure at menarche, but not at later stages of life.[38]

We conducted additional analyses to assess the influence of certain factors on our findings. When performing analyses with a 5- and 10-year exposure lag, similar patterns were found compared to the main analyses, which indicated that recent exposure was not influential. Furthermore, similar findings were observed when analyzing only confirmed cases, which discounted the influence of outcome misclassification among the “high probability” cases. Inclusion of “high probability” cases alongside confirmed cases mitigated potential bias from excluding cases who did not have resources or were unwilling or otherwise unable to undergo additional diagnostic procedures.

Our study had notable strengths. First, we had high participation rates and likely captured most incident lung cancer cases in Xuanwei and Fuyuan. Second, we used geospatial models to predict coal deposit use for participants across their lifecourse. Previous geochemical analyses of coal composition found that a considerable proportion of residents (~10%) in Xuanwei misreported the type of coal that was used.[15] As such, prediction of coal deposit use based on residential location and other household factors may provide greater accuracy than self-report in this rural population. Third, the study population was composed of Chinese women who never smoked, which precludes confounding by active smoking, Ethnic group/ethnicity, and sex. Fourth, differential misclassification of the exposure was improbable, given that deposit use was predicted from residential and coal mine location, which were unlikely to be reported differently between cases and controls.

There were several potential limitations of our study. Due to sparse data for certain comparisons, some of the estimates may be imprecise. However, the findings for smoky coal deposits in this region were consistent with those that were mapped to particular administrative regions in an earlier population-based case–control study conducted in the 1980s[13] and with results from a previous retrospective cohort study that broadly collapsed coal into only two types (i.e., smoky and smokeless).[7] This finding also provides support that several potentially important biases from the inherent design of a case–control study did not substantially impact our results. Further, the sparse data from specific deposits required combining several deposits; therefore, finer spatial variation in risks was not possible. Additionally, deposit subtype classifications were based on relatively broad criteria. While sufficient for identifying variation in risks and constitutional makeup, these criteria made finer scale examination of coal strata infeasible. However, this finer degree of examination by geological source and strata may not be relevant to epidemiological studies, as multiple strata exist within each deposit, and coal mines collect whichever coal is available regardless of individual strata.

In summary, this was the first study to our knowledge to assess the variation in lung cancer risk by geologic source deposit of household coal. The substantially elevated risks of lung cancer from use of smoky coal in Xuanwei and Fuyuan differed according to geologic source deposit, with those in Laibin township being among the most detrimental. Further, smoky coal deposits in Fuyuan were associated with elevated risks comparable to those of Laibin, which are historically among the highest in China, suggesting that there may be additional high-risk areas of China that have yet to be identified. Additionally, childhood exposure to smoky coal may be important in influencing lung cancer risk later in life. Taken together, our findings suggest that the health consequences attributed to early-life and frequent exposure to certain types of combustion emissions may vary in magnitude according the geologic source deposit of coal, and that these factors should be considered when evaluating the health risk posed by exposure to coal combustion. Our study lays the foundation for future investigations of coal combustion constituents that attempt to identify and characterize the specific components that drive geographic variation in health risks.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

None of the authors declare any actual or potential competing financial interests. This work was supported with intramural funding from the National Cancer Institute. We thank Jackie King, Peter Hui, Nathan Appel, and Leslie Carroll for their support. We dedicate this work to the memory of Robert Chapman and Xingzhou He, outstanding mentors and pioneers in the study of the lung cancer excess in this part of the world.

APPENDIX

Assignment of coal deposit was based primarily on imputation that was performed using a Bayesian approach. Briefly, a multinomial choice model was created for each village in the dataset that included all mines within a 30 km radius of the village center and separately for subjects reporting either using smoky, smokeless, or mixed coal sources. Prior probabilities for each mine were derived in a way that resulted in 50% lower odds of selection for each additional 10 km increase in distance from the village center, and in five-fold higher odds of selection for mines that produced the coal type that was reported for subjects reporting either smoky or smokeless (not mixed) coal use. The rationale for these choices was to allow a higher probability of selection for mines that were closer to the village where a participant lived and that produced the same coal as the subject reported using, while still allowing for potential reporting errors.

Selection probabilities for each deposit were calculated by summing up selection probabilities for individual mines that were mining that deposit, and the five most likely deposits were retained for use in prediction of exposure parameters and sensitivity analysis. The current analyses used the most likely predicted deposit for each participant that reported using coal as the primary fuel source.

Imputation was based on the multinomial choice model, and separate models were fitted for each village and separately for participants who reported use of smoky, smokeless, or mixed coal sources. A dataset containing information on self-reported mine use was assembled for each village and coal use stratum, by combining information from the present case–control study, an older case–control study,[13] a cohort study,[7] and an exposure assessment study[11,14,15] that were conducted in the same region.

To reduce the dimensionality of the models, the number of mines under consideration for each village was limited to those mines that were within a 30 km radius of the village center, which was the upper 90% quantile of observed distances in our dataset. This included mines that were not chosen by any of the participants for which the coal source was known (i.e., for which the observed counts were 0). The number of mines under consideration for each village ranged from 35 to 864, with a median of 578. The dataset included a total of 13,112 observed mine choices, and the number of observed choices for subjects that needed imputation ranged from 0 to 738, with a median of 13.

The following presentation largely follows that in Agresti and Hitchcock.[39] With N subjects which have a known coal source and M potential mines to choose from, we assume that the observed choices $m = (m_1, m_2, \ldots, m_M)$ follow a multinomial distribution with $\Sigma m_i = N$ and parameters (multinomial probabilities) $\pi = (\pi_1, \pi_2, \ldots, \pi_m)$. A maximum-likelihood based analysis would result in estimates for π equal to the observed proportions

$(m_1/N, m_2/N, \ldots, m_m/N)$, and would result in zero probabilities for all choices that were not actually observed, even though the number of observed choices (i.e., N) may be quite low compared to the number of mines to choose from (i.e., M) as we have here. The Bayesian analysis that we performed instead allows us to weigh this likelihood-based information with any other prior information on mine choice probabilities that might be available.

In a Bayesian analysis, the conjugate prior for the multinomial likelihood is the Dirichlet distribution with hyperparameter α, and $\alpha_i > 0$. Let $K = \Sigma\alpha_i$. Then the Dirichlet has $E(\pi_i) = \alpha_i/K$, and the posterior mean $E(\pi_i \mid m_1, m_2, \ \ldots, \ m_M) = (m_i + \alpha_i)/(M + K)$. By allowing the α_i's to be different for different mines we can allow different *a-priori* probabilities of selection, and by chancing the absolute value of the α_i's (and thus K) we can affect the relative weights of the prior and observed probabilities (mine choices) in producing the posterior probability of selection for each mine.

We chose the α_i to favor correspondence between the self-reported coal type and the coal type produced by any imputed mines, by fitting separate model for subjects that report using either smoky, smokeless, or mixed coals, and by setting α_i for those mines that produce the self-reported type of coal to be 5 times that for the other mines (or equal for the stratum of subjects reporting mixed coal use). We further allowed the mine-specific *a priori* probabilities to depend on the distance of the mine to the village center, with mines that were further away having lower selection probabilities. Preliminary analyses of the available data suggested that the probability of selection approximately halved for each additional 10 km in distance, so we modified the α_i's accordingly. Finally, we followed the suggestion by Perks (1947) and let $K = \Sigma\alpha_i = 1$.

Models were fitted in JAGS 3.4.0 using the rjags package in R for processing of the results.

Abbreviations:

CI	95% confidence intervals
FY	Fuyuan
IARC	The International Agency for Research on Cancer
ICD-9	International Classification of Disease 9th revision
OR	odds ratios
SES	socioeconomic status
TB	tuberculosis
XW	Xuanwei

References

1. World Health Organization (WHO). Air Pollution, percentage of population using biomass fuels. Millennium Indicators Database, United Nations, Department of Economic and Social

Affairs, Economic and Social Development, Statistics Division. https://millenniumindicators.un.org/unsd/mdg/

2. International Agency for Research on Cancer Monograph Working Group I. 2013;103. https://monographs.iarc.fr/iarc-monographs-on-the-evaluation-of-carcinogenic-risks-to-humans-17/
3. International Agency for Research on Cancer Monograph Working Group I. 2005;92. https://monographs.iarc.fr/iarc-monographs-on-the-evaluation-of-carcinogenic-risks-to-humans-17/
4. International Agency for Research on Cancer Monograph Working Group I.100E. https://monographs.iarc.fr/iarc-monographs-on-the-evaluation-of-carcinogenic-risks-to-humans-17/
5. Hosgood HD 3rd, Wei H, Sapkota A, et al. Household coal use and lung cancer: systematic review and meta-analysis of case-control studies, with an emphasis on geographic variation. Int J Epidemiol 2011;40:719–28. [PubMed: 21278196]
6. Data records of the cancer mortality survey of China in 1973–1975. The People's Health Publishing House, 1980.
7. Barone-Adesi F, Chapman RS, Silverman DT, et al. Risk of lung cancer associated with domestic use of coal in Xuanwei, China: retrospective cohort study. BMJ 2012;345:e5414. [PubMed: 22936785]
8. Lan Q, Chapman RS, Schreinemachers DM, et al. Household stove improvement and risk of lung cancer in Xuanwei, China. J Natl Cancer Inst 2002;94:826–35. [PubMed: 12048270]
9. Mumford JL, He XZ, Chapman RS, et al. Lung cancer and indoor air pollution in Xuan Wei, China. Science 1987;235:217–20. [PubMed: 3798109]
10. Li XM. Carcinogenicity and organic fraction of indoor inhalable particles. Chin J Prev Med 1989;23:358–60.
11. Hu W, Downward GS, Reiss B, et al. Personal and indoor PM2.5 exposure from burning solid fuels in vented and unvented stoves in a rural region of China with a high incidence of lung cancer. Environ Sci Technol 2014;48:8456–64. [PubMed: 25003800]
12. Kim C, Chapman RS, Hu W, et al. Smoky coal, tobacco smoking, and lung cancer risk in Xuanwei, China. Lung Cancer 2014;84:31–5. [PubMed: 24506909]
13. Lan Q, He X, Shen M, et al. Variation in lung cancer risk by smoky coal subtype in Xuanwei, China. Int J Cancer 2008;123:2164–9. [PubMed: 18712724]
14. Downward GS, Hu W, Rothman N, et al. Polycyclic aromatic hydrocarbon exposure in household air pollution from solid fuel combustion among the female population of Xuanwei and Fuyuan counties, China. Environ Sci Technol 2014;48:14632–41. [PubMed: 25393345]
15. Downward GS, Hu W, Large D, et al. Heterogeneity in coal composition and implications for lung cancer risk in Xuanwei and Fuyuan counties, China. Environ Int 2014;68:94–104. [PubMed: 24721117]
16. Finkelman RB. Health impacts of coal: facts and fallacies. Ambio 2007;36:103–6. [PubMed: 17408201]
17. Finkelman RB, Belkin HE, Zheng B. Health impacts of domestic coal use in China. Proc Natl Acad Sci USA 1999;96:3427–31. [PubMed: 10097053]
18. Large DJ, Kelly S, Spiro B, et al. Silica-volatile interaction and the geological cause of the Xuan Wei lung cancer epidemic. Environ Sci Technol 2009;43:9016–21. [PubMed: 19943682]
19. Chen P Study on integrated classification system for Chinese coal. Fuel Process Technol 2000;62:77–87.
20. Rothman KJ, Greenland S, Lash TL. Modern Epidemiology, 3rd ed. Philadelphia, PA: Wolters Kluwer Health/Lippincott Williams & Wilkins, 2008. x, 758 p.
21. Samet JM, Humble CG, Pathak DR. Personal and family history of respiratory disease and lung cancer risk. Am Rev Respir Dis 1986;134:466–70. [PubMed: 3752703]
22. Denholm R, Schuz J, Straif K, et al. Is previous respiratory disease a risk factor for lung cancer? Am J Respir Crit Care Med 2014;190:549–59. [PubMed: 25054566]
23. Clarke MA, Joshu CE. Early life exposures and adult cancer risk. Epidemiol Rev 2017;39:11–27. [PubMed: 28407101]
24. Barton HA, Cogliano VJ, Flowers L, et al. Assessing susceptibility from early-life exposure to carcinogens. Environ Health Perspect 2005;113:1125–33. [PubMed: 16140616]

25. Zhang JJ, Smith KR. Household air pollution from coal and biomass fuels in China: measurements, health impacts, and interventions. Environ Health Perspect 2007;115:848–55. [PubMed: 17589590]
26. Mumford JL, Li X, Hu F, et al. Human exposure and dosimetry of polycyclic aromatic hydrocarbons in urine from Xuan Wei, China with high lung cancer mortality associated with exposure to unvented coal smoke. Carcinogenesis 1995;16:3031–6. [PubMed: 8603481]
27. Mumford JL, Lee X, Lewtas J, et al. DNA adducts as biomarkers for assessing exposure to polycyclic aromatic hydrocarbons in tissues from Xuan Wei women with high exposure to coal combustion emissions and high lung cancer mortality. Environ Health Perspect 1993;99:83–7. [PubMed: 8319664]
28. Wong JYY, Hu W, Downward GS, et al. Personal exposure to fine particulate matter and benzo[a] pyrene from indoor air pollution and leukocyte mitochondrial DNA copy number in rural China. Carcinogenesis 2017;38:893–9. [PubMed: 28911003]
29. Seow WJ, Downward GS, Wei H, et al. Indoor concentrations of nitrogen dioxide and sulfur dioxide from burning solid fuels for cooking and heating in Yunnan Province, China. Indoor Air 2016;26:776–83. [PubMed: 26340585]
30. Knudson AG Jr. Mutation and cancer: statistical study of retinoblastoma. Proc Natl Acad Sci USA 1971;68:820–3. [PubMed: 5279523]
31. Clark JD 3rd, Serdar B, Lee DJ, et al. Exposure to polycyclic aromatic hydrocarbons and serum inflammatory markers of cardiovascular disease. Environ Res 2012;117:132–7. [PubMed: 22626472]
32. Waidyanatha S, Zheng Y, Serdar B, et al. Albumin adducts of naphthalene metabolites as biomarkers of exposure to polycyclic aromatic hydrocarbons. Cancer Epidemiol Biomarkers Prev 2004;13:117–24. [PubMed: 14744742]
33. Li J, Zhu X, Yu K, et al. Exposure to polycyclic aromatic hydrocarbons and accelerated DNA methylation aging. Environ Health Perspect 2018;126:067005. [PubMed: 29906262]
34. Bijnens EM, Zeegers MP, Derom C, et al. Telomere tracking from birth to adulthood and residential traffic exposure. BMC Med 2017;15:205. [PubMed: 29157235]
35. Schultz ES, Litonjua AA, Melen E. Effects of long-term exposure to traffic-related air pollution on lung function in children. Curr Allergy Asthma Rep 2017;17:41. [PubMed: 28551888]
36. Gehring U, Gruzieva O, Agius RM, et al. Air pollution exposure and lung function in children: the ESCAPE project. Environ Health Perspect 2013;121:1357–64. [PubMed: 24076757]
37. Oftedal B, Brunekreef B, Nystad W, et al. Residential outdoor air pollution and lung function in schoolchildren. Epidemiology 2008;19:129–37. [PubMed: 18091005]
38. Nie J, Beyea J, Bonner MR, et al. Exposure to traffic emissions throughout life and risk of breast cancer: the Western New York exposures and breast cancer (WEB) study. Cancer Causes Control 2007;18:947–55. [PubMed: 17632764]
39. Agresti A, Hitchcock D. Bayesian inference for categorical data analysis. Stat Methods Appl 2005;14:297–330.

What's new?

Indoor coal combustion emissions have been designated as Group 1 carcinogens. However, the influence of emissions from coal originating from different geological deposits remains unexplored. Lung cancer rates in the rural counties of Xuanwei and Fuyuan in China are among the highest in the world due to indoor combustion of bituminous ("smoky") coal. Using detailed lifecourse data on household fuel sources, the authors conducted the first study to show that lung cancer risk varies by the geological deposit of household coal. Notably, childhood exposure to smoky coal may have an important impact on lung cancer risk later in life.

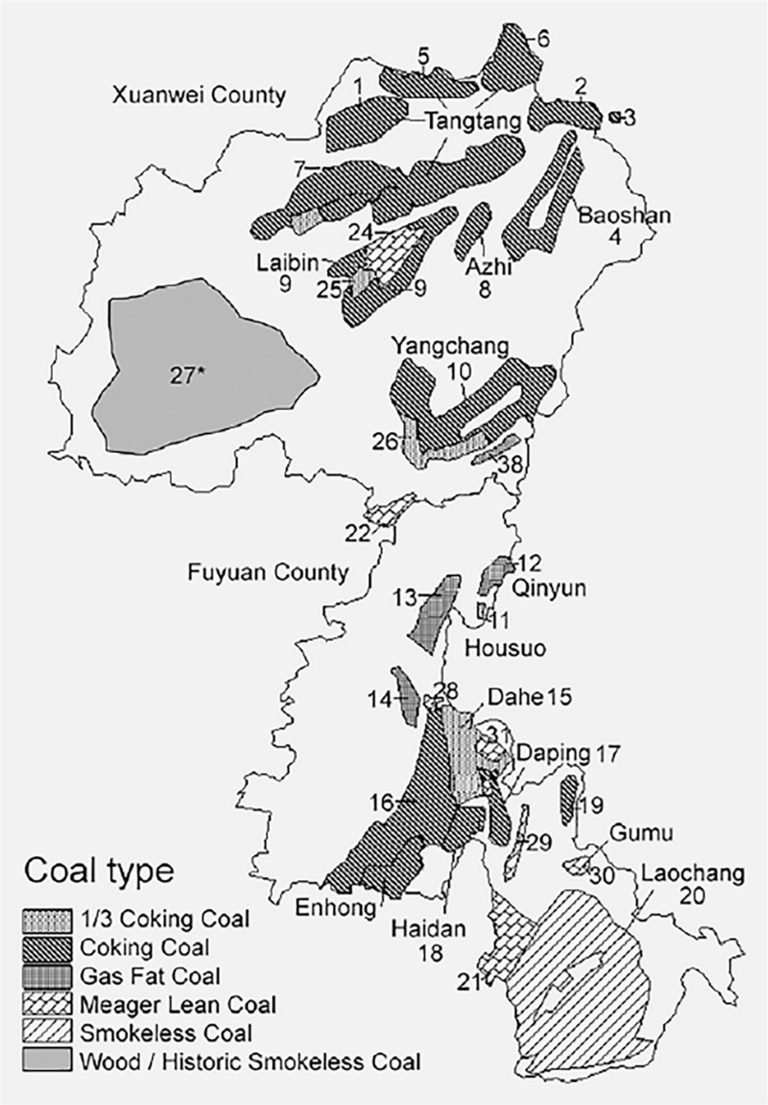

Figure 1.

Map of Geological Coal Deposits in Xuanwei and Fuyuan Counties of Yunnan, China. Coal deposits are designated by numbers. The distinctiveness of geologic deposits across Xuanwei and Fuyuan, which reflect differences in coalification or depositional environment, are indicated by the subcategorization of smoky coal into coking coal, 1/3 coking coal, gas fat coal and meager lean coal. This subcategorization was based on the degree of coalification and caking property characteristics as defined by the State Standard of China Coal Classification. *Historically a smokeless coal area that is recently transitioning to smoky coal, wood, and dung.

Table 1.

Study population characteristics of female never-smoking lung cancer cases and population-based controls in Xuanwei and Fuyuan, China

	Cases		Controls		*p* Value
Characteristic	***n* = 1031**		***n* = 493**		
Age, years, mean, SD	54.7	10.6	54.7	11.4	1.00
County, *n*, %					
Xuanwei	641	62.2	296	60.0	0.42
Fuyuan	390	37.8	197	40.0	
Ever used smoky coal, *n*, %					
No	21	2.0	59	12.0	7.0E-15[1]
Yes	1006	97.6	432	87.6	
Lifetime smoky coal use, tons, mean, SD	178	101	148	121	7.8E-12[1]
Ever exposed to secondhand smoke, *n*, %					
No	41	4.0	10	2.0	4.9E-02[1]
Yes	990	96.0	483	98.0	
Ever used a stove with a chimney (ventilation)					
No	100	9.7	60	12.1	0.16
Yes	917	88.9	431	87.4	
Food sufficiency before marriage (socioeconomic status)					
More than enough	116	11.3	30	6.1	9.5E-03[1]
Just enough	335	32.5	158	32.0	
Not enough	574	55.7	303	61.5	
Education, n, %					
Illiterate/No school	642	62.3	286	58.0	2.3E-03[1]
Attended elementary school	171	16.6	117	23.7	
Graduated elementary school	102	9.9	32	6.5	
Attended middle school or higher	115	11.2	57	11.6	
First degree relative with cancer, n, %					
No	821	79.6	459	93.1	9.7E-13[1]
Yes	205	19.9	32	6.5	

	Cases		Controls		*p* Value
Characteristic	***n* = 1031**		***n* = 493**		
History of respiratory disease (asthma, bronchitis, emphysema, tuberculosis), n, %					
No	927	89.9	473	95.9	5.6E-05[1]
Yes	104	10.1	20	4.1	

[1] *p*-Values <0.05 were considered statistically significant. Continuous variables were compared using Wilcoxon rank sum tests. Categorical variables were compared using Chi-square or Fisher's Exact tests.

Discrepancy in counts due to missing data.

Table 2.

Associations between coal deposit and lung cancer risk in a population-based case-control study of female never-smokers from Xuanwei and Fuyuan, China

Coal deposit	Coal region	Coal subcategorization	Cases	Controls	OR	95% CI Low	95% CI Up	p Value
I) Deposit used most frequently throughout lifecourse								
Never smoky coal users (Reference)	XW and FY	Wood/dung/Smokeless coal users from Smokeless coal regions	13	38	1.00			
Combined Smoky Coal Deposits	XW and FY		844	207	12.57	6.51	24.24	4.3E-14
9	XW, Laibin	Coking	101	9	33.40	13.07	85.34	2.3E-13
24	XW, Laibin	Meager Lean	43	6	20.86	7.16	60.78	2.6E-08
1,2,4,7,8	XW, North	Coking	336	100	10.38	5.28	20.42	1.2E-11
10	XW South	Coking	63	27	7.49	3.43	16.38	4.6E-07
12,13,14,38[1]	FY, North	Gas Fat	48	9	16.32	6.26	42.54	1.1E-08
16,17,19	FY, South	Coking	253	56	14.27	7.07	28.78	1.2E-13
20, 27	XW and FY	Ever smoky coal users who lived in smokeless coal regions	8	5	5.09	1.40	18.43	0.01
Wood/dung users	XW and FY	-	39	59	2.01	0.94	4.28	0.07
Other/Outside/Unknown	Undefined	-	127	184	2.07	1.05	4.07	0.04
II) Deposit used at birth home								
Never smoky coal users (Reference)	XW and FY	Wood/dung/Smokeless coal users from Smokeless coal regions	13	38	1.00			
Combined Smoky Coal Deposits	XW and FY		839	167	15.68	8.10	30.35	3.0E-16
9	XW, Laibin	Coking	117	15	24.15	10.45	55.81	9.5E-14
24	XW, Laibin	Meager Lean	33	4	24.08	7.08	81.89	3.5E-07
1,2,4,7,8	XW, North	Coking	319	75	13.21	6.65	26.23	1.7E-13
10	XW South	Coking	58	22	8.56	3.82	19.17	1.8E-07
12,13,14,38[1]	FY, North	Gas Fat	54	5	34.08	11.14	104.26	6.2E-10
16,17,19	FY, South	Coking	258	46	17.81	8.74	36.29	2.2E-15
20, 27	XW and FY	Ever smoky coal users who lived in smokeless coal regions	5	2	8.09	1.39	47.09	0.02
Wood/dung users	XW and FY	-	65	115	1.71	0.85	3.48	0.14
Other/Outside/Unknown	Undefined	-	109	171	1.97	1.00	3.89	0.05

p-Values <0.05 were considered statistically significant. Logistic regression models were adjusted for age and food sufficiency before marriage or age 20 years (surrogate for socioeconomic status).

[1] Deposit 38 is located in southern Xuanwei.

Table 3.

Coal deposit and lifetime tonnage on lung cancer risk in a population-based case-control study of female never-smokers from Xuanwei and Fuyuan, China

Coal deposit	Coal lifetime tonnage	Coal region/subcategorization	Cases	Controls	OR	95% CI Low	95% CI Up	*p* Value
I) Deposit used most frequently throughout lifecourse								
Never smoky coal users (Reference)	-	Wood/dung/Smokeless coal users from Smokeless coal regions	13	38	1.00			
9	Low	XW, Laibin, Coking	30	5	16.85	5.33	53.33	1.5E-06
9	High	XW, Laibin, Coking	68	4	48.75	14.77	160.89	1.8E-10
24	Low	XW, Laibin, Meager Lean	14	2	18.07	3.57	91.57	4.7E-04
24	High	XW, Laibin, Meager Lean	27	4	20.28	5.93	69.35	1.6E-06
1,2,4,7,8	Low	XW, North, Coking	64	27	7.09	3.23	15.58	1.1E-06
1,2,4,7,8	High	XW, North, Coking	241	73	10.00	5.02	19.89	5.4E-11
10	Low	XW, South, Coking	13	11	3.52	1.25	9.90	0.02
10	High	XW, South, Coking	42	16	8.38	3.55	19.79	1.2E-06
12,13,14,38[1]	Low	FY, North, Gas Fat	6	2	9.61	1.71	54.21	0.01
12,13,14,38[1]	High	FY, North, Gas Fat	40	7	16.86	6.05	47.03	6.7E-08
16,17,19	Low	FY, North, Coking	56	19	8.84	3.85	20.31	2.8E-07
16,17,19	High	FY, North, Coking	188	37	15.65	7.56	32.40	1.3E-13
20, 27	-	Ever smoky coal users who lived in smokeless coal regions in XW and FY	8	5	4.99	1.38	18.06	0.01
Wood/dung users	-	XW and FY	39	59	1.98	0.93	4.22	7.5E-02
Other/Outside/Unknown	-	-	182	184	2.92	1.50	5.70	1.7E-03
II) Deposit used at birth home								
Never smoky coal users (Reference)	-	Wood/dung/Smokeless coal users from Smokeless coal regions	13	38	1.00			
9	Low	XW, Laibin, Coking	32	8	11.54	4.18	31.86	2.3E-06
9	High	XW, Laibin, Coking	76	7	32.15	11.79	87.64	1.2E-11
24	Low	XW, Laibin, Meager Lean	8	2	10.11	1.85	55.29	7.6E-03
24	High	XW, Laibin, Meager Lean	24	2	34.39	7.09	166.82	1.1E-05
1,2,4,7,8	Low	XW, North, Coking	71	20	10.92	4.83	24.65	8.9E-09

Coal deposit	**Coal lifetime tonnage**	**Coal region/subcategorization**	**Cases**	**Controls**	**OR**	**95% CI Low**	**95% CI Up**	***p* Value**
1,2,4,7,8	High	XW, North, Coking	228	55	12.44	6.18	25.06	1.7E-12
10	Low	XW, South, Coking	12	11	3.46	1.21	9.84	0.02
10	High	XW, South, Coking	40	11	11.49	4.57	28.90	2.1E-07
12,13,14,38[1]	Low	FY, North, Gas Fat	8	2	13.67	2.54	73.58	2.3E-03
12,13,14,38[1]	High	FY, North, Gas Fat	45	3	44.72	11.82	169.21	2.2E-08
16,17,19	Low	FY, North, Coking	56	14	12.49	5.20	30.01	1.7E-08
16,17,19	High	FY, North, Coking	191	32	18.28	8.74	38.21	1.1E-14
20, 27	-	Ever smoky coal users who lived in smokeless coal regions in XW and FY	5	2	7.95	1.37	46.21	0.02
Wood/dung users	-	XW and FY	65	115	1.69	0.84	3.43	0.14
Other/Outside/Unknown	-	-	157	171	2.77	1.42	5.43	3.0E-03

p-Values <0.05 were considered statistically significant. Lifetime smoky coal tonnage was dichotomized by the median among the controls (Low: <125 High: ≥125). Logistic regression models were adjusted for age and food sufficiency before marriage or age 20 years (surrogate for socioeconomic status).

[1] Deposit 38 is located in southern Xuanwei.

Table 4.

Associations between change in fuel use and lung cancer risk in a population-based case-control study of female never-smokers from Xuanwei and Fuyuan, China

Fuel use	Cases	Controls	OR	95% CI Low	95% CI Up	*p* Value
Never smoky coal users who lived in smokeless coal areas and ever used smokeless coal, wood, dung, gas, or electricity (reference)	13	38	1.00			
Lifetime smoky coal users	603	89	19.64	9.95	38.77	9.2E-18
Smoky coal users at birth home who changed to smokeless coal, wood, dung, gas or electricity permanently	63	15	10.83	4.61	25.46	4.6E-08[1]
Smokeless coal, wood, dung, gas or electricity users at birth home who changed to smoky coal permanently	33	22	4.74	2.03	11.04	3.2E-04[1]
Other	319	329	2.75	1.42	5.31	2.6E-03

p-Values <0.05 were considered statistically significant. Logistic regression model was adjusted for age and food sufficiency before marriage or age 20 years (surrogate for socioeconomic status) and presence of ventilation.

[1] $p = 0.04$ for difference, determined by using the category “Smoky coal users at birth who changed to smokeless coal, wood, dung, gas or electricity permanently” as the referent.