ORIGINAL ARTICLE

Determinants of overall survival in lung cancer in Kazakhstan: a retrospective cohort study using classical survival analysis

Zeinep Avizova[1,2], Ayan Myssayev[3,4], Yerbolat Iztleuov[1], Nadiar Mussin[5], Amin Tamadon[6]

[1]Department of Radiology, West Kazakhstan Marat Ospanov Medical University, Aktobe, Republic of Kazakhstan; [2]Department of Hygiene and Epidemiology, South Kazakhstan Medical Academy, Shymkent, Republic of Kazakhstan; [3]Department of Science and Human Resources, Ministry of Health of the Republic of Kazakhstan, Astana, Republic of Kazakhstan; [4]Office of the Vice-Rector for Strategic Development and International Cooperation, Astana Medical University, Astana, Kazakhstan; [5]Department of General Surgery, West Kazakhstan Marat Ospanov Medical University, Aktobe, Republic of Kazakhstan; [6]Department of Natural Sciences, West Kazakhstan Marat Ospanov Medical University, Aktobe, Republic of Kazakhstan.

ABSTRACT

Background: Lung cancer is the leading cause of cancer-related death worldwide, but population-based survival data from Central Asia are scarce. Kazakhstan has a substantial lung cancer burden, yet little is known about survival and its determinants. This study aimed to identify demographic and clinical predictors of overall survival (OS) among lung cancer patients in Kazakhstan.
Research design and Methods: We conducted a retrospective cohort study using the National Cancer Registry of Kazakhstan. Adults (≥18 years) with primary lung cancer diagnosed between 1 January 2019 and 31 December 2023 were analyzed. OS was defined as days from diagnosis to death or last follow-up. Kaplan–Meier methods estimated OS and compared survival by demographic, socioeconomic, and tumor characteristics. Cox proportional hazards models identified independent prognostic factors.
Results: 13,402 patients were included (mean age 64.1 years; 79.4% male; 62.2% urban). Squamous cell carcinoma (34.9%) and stage III disease (47.5%) predominated. Median OS was 167 days (IQR 53–399), and 75.9% of patients died during follow-up. In multivariable analyses, older age, male sex, lower social status, unfavorable diagnostic circumstances, non-adenocarcinoma histology, and advanced stage were associated with higher mortality, while urban residence was protective. Tumor stage was the strongest predictor of OS.

Received: 5 December 2025 | Accepted: 10 January 2026
Correspondence: Zeinep Avizova, MSc / Department of Radiology, West Kazakhstan Marat Ospanov Medical University, Aktobe, Kazakhstan / E-mail: zeinepavizova@gmail.com
ORCID: 0000-0001-5540-0229

Conclusions: Lung cancer survival in Kazakhstan is poor, with patients presenting at advanced stages. Demographic and socioeconomic characteristics are independent determinants of OS. These findings support efforts to improve early detection, broaden access to lung cancer care—especially for rural and disadvantaged populations—and incorporate risk stratification into cancer control policies.

Key words: lung neoplasms, survival analysis, proportional hazards models, socioeconomic factors, rural population

Introduction

Lung cancer is the leading cause of cancer-related mortality worldwide, accounting for approximately 1.8 million deaths annually (1). Its high burden is driven by a combination of risk factors, including tobacco use, environmental exposures, and genetic predispositions, with significant variations in incidence and outcomes across regions (2). In Kazakhstan, lung cancer poses a substantial public health challenge, yet comprehensive population-based studies on its clinicopathological characteristics and survival outcomes are scarce. Previous research in Kazakhstan has indicated a higher incidence among males and regional disparities, but detailed analyses of prognostic factors and survival patterns are limited (3). The heterogeneity of lung cancer, characterized by diverse histological subtypes (e.g., adenocarcinoma, squamous cell carcinoma, small cell carcinoma) and varying stages at diagnosis, complicates prognosis and treatment strategies (4). Global trends suggest that early detection through screening, precise histological classification, and tailored therapies improve survival, yet these interventions are underutilized in many low- and middle-income countries, including Kazakhstan (5). Additionally, socioeconomic factors, such as rural-urban disparities and social status, may influence access to care and outcomes, warranting investigation in the Kazakhstani context (6). This study aims to address these gaps by analyzing data from the National Cancer Registry of Kazakhstan (2019–2023) to characterize the clinicopathological features and survival outcomes of lung cancer patients. By examining demographic and clinical predictors of overall survival (OS) and identifying demographic and clinicopathological predictors of overall survival, we seek to provide insights into prognostic factors specific to this population. These findings could inform targeted interventions, such as enhanced screening programs and improved healthcare access, to reduce the lung cancer burden in Kazakhstan.

Materials and methods

Study design, data source, and study population

This retrospective cohort study analyzed data from the National Cancer Registry of Kazakhstan, maintained by the Republican Center for Healthcare Development under the Ministry of Health of the Republic of Kazakhstan. The registry provides anonymized, population-based data on all cancer cases nationwide. The study included all adults (≥18 years) newly diagnosed with primary lung cancer between January 1, 2019, and December 31, 2023. Cases were eligible if they had complete information on age, sex, nationality, place of residence, social status, diagnostic circumstances, tumor histology, disease stage, survival status, and survival time. Patients with missing survival data or secondary lung malignancies were excluded. After applying these criteria, 13,402 patients were included in the final analysis (Figure 1).

Data collection procedures

Data were abstracted by trained registry personnel from hospital medical records, pathology reports,

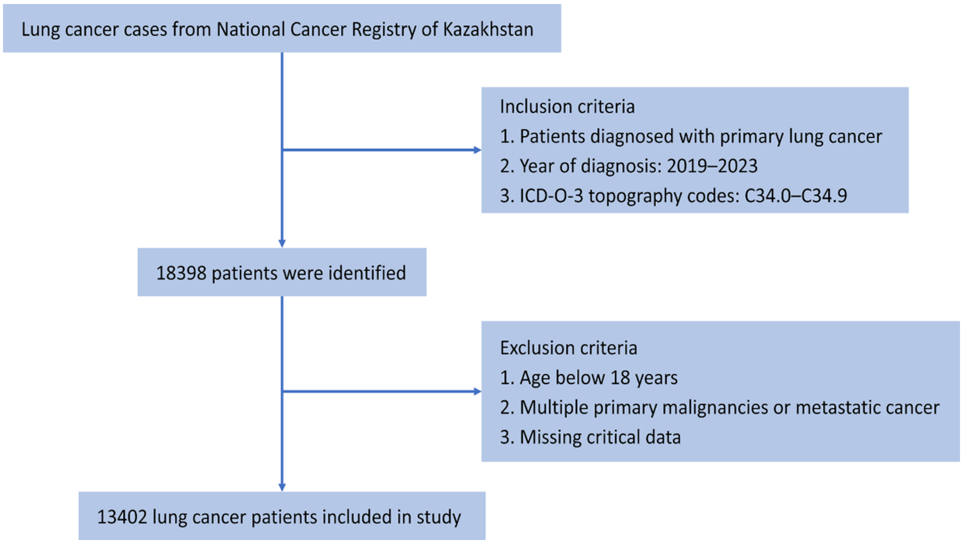


Figure 1. Flowchart of study population selection for clinicopathological characteristics and survival outcomes in lung cancer patients in Kazakhstan from the national cancer registry of Kazakhstan (2019–2023).

and death certificates using standardized procedures. Recorded variables included demographic characteristics (age, sex, nationality, residence, social status), diagnostic circumstances (coded as categories 1–3), tumor histology, and stage at diagnosis. Survival status and dates of diagnosis and death were obtained from national mortality records. All data were anonymized prior to analysis.

Variables and definitions

Age was analyzed both as a continuous variable and an ordinal four-level age-group variable (1–4). Sex was categorized as male or female. Nationality was coded in three categories. Residence was classified as urban or rural. Social status was recorded using three ordinal categories, coded as 1 (high), 2 (intermediate), and 3 (low) social status. Diagnostic circumstances were grouped into three categories (codes 1–3). Tumor histology consisted of six mutually exclusive categories: adenocarcinoma, squamous cell carcinoma, small-cell carcinoma, large-cell carcinoma, NSCLC otherwise specified, and other malignancies. Disease stage was categorized according to TNM staging as Stage I–IV. Overall survival (OS) was defined as the number of days from diagnosis to death from any cause or last known follow-up. Patients alive at the end of follow-up were censored.

Data management and preprocessing

Data cleaning included verification of valid ranges, exclusion of patients with survival time ≤0 days, and recoding of categorical variables into interpretable labels. No missing values were present in the variables used for survival analysis after preprocessing.

Quality control measures

Registry data undergo routine validation and cross-checking against primary medical documentation. Histological and staging discrepancies were resolved by reference to original pathology reports. Mortality data were validated using national vital statistics records to minimize loss to follow-up.

Ethical considerations

The study received approval from the Bioethics Committee of West Kazakhstan Medical University (Protocol No. 3, March 28, 2025). As anonymized registry data were used, informed consent was not required.

Survival analysis by demographic and clinicopathological characteristics

Overall survival was estimated using Kaplan–Meier methods and stratified by age, sex, nationality, residence, social status, diagnostic circumstances, histology, and tumor stage.

Statistical analysis

All statistical analyses were conducted using Python (version 3.10) with the lifelines, pandas, numpy, and matplotlib libraries. Baseline demographic and clinicopathological characteristics were summarized using means and standard deviations for continuous variables and frequencies with percentages for categorical variables; age was analyzed both as a continuous variable and as an ordinal four-level variable. Overall survival (OS) was defined as the number of days from diagnosis to death from any cause or to the last follow-up, with censored observations assigned to patients who were alive at the end of the study period. OS was estimated using the Kaplan–Meier method, and survival curves were stratified according to age, sex, nationality, place of residence, social status, diagnostic circumstances, histological subtype, and tumor stage. Differences between survival curves were evaluated using the log-rank (Mantel–Cox) test, and median survival times with corresponding 95% confidence intervals (CIs) were reported for each subgroup. To identify prognostic factors for overall survival, univariate Cox proportional hazards regression models were first fitted for each variable. Variables demonstrating clinical relevance or statistical significance in univariate analyses were subsequently included in the multivariable Cox regression model. The final multivariable model included age, sex, nationality, place of residence, social status, diagnostic circumstances, histological subtype, and tumor stage, and adjusted hazard ratios (HRs) with 95% CIs were estimated for each covariate. Model diagnostics were performed to evaluate the adequacy of the Cox model. The proportional hazards (PH) assumption was assessed using scaled Schoenfeld residuals and global significance tests, supplemented by visual inspection of covariate-specific residual plots. Linearity of the continuous predictor age was examined using martingale residual plots, with no major deviations from linearity observed. Multicollinearity among covariates was assessed using variance inflation factors (VIF), with all VIF values falling below commonly accepted thresholds, indicating the absence of problematic collinearity. Influence diagnostics were conducted using delta-beta ($\Delta\beta$) residuals to detect observations exerting disproportionate influence on model coefficients; no significant outliers were identified. These diagnostic procedures collectively supported the stability and validity of the final Cox model. All statistical tests were two-sided, and a p-value <0.05 was considered statistically significant. Hazard ratios were reported with corresponding 95% confidence intervals.

Results

Baseline characteristics

A total of 13,402 patients with lung cancer were included in this study. The mean age at diagnosis was 64.1 years (SD = 9.5; range 18–94). Most patients were male (79.4%), and the majority resided in urban areas (62.2%). The dominant nationality category was Kazakh followed by Russian and others. Social status was distributed with a mean level of 1.92 (SD = 0.75), indicating that, on average, patients clustered between intermediate and low social status categories. Most patients were diagnosed under symptomatic circumstances (DiagCircumstance = 2 was the most common, mean = 1.74). Regarding histopathology, squamous cell carcinoma was the most frequent subtype (34.9%), followed by adenocarcinoma and other NSCLC variants. In terms of stage distribution, the most frequent was Stage III (47.5%), followed by Stage I, II, and IV. The median overall survival time was 167 days (IQR 53–399), with a range of 1 to 1823 days. At the end of follow-up, 75.9% of patients were deceased (OS = 1), and 46.2% had cancer-specific mortality (CSS = 1). Baseline characteristics are summarized in Table 1.

Kaplan–Meier survival curves

Kaplan–Meier survival analyses demonstrated significant differences in overall survival across key clinical and sociodemographic subgroups.

Survival by tumor stage

There was a clear, stepwise decline in overall survival with advancing tumor stage (Figure 2). Patients diagnosed with Stage I disease experienced the most favorable survival, followed by progressively poorer outcomes in Stages II, III, and IV. The differences across stages were statistically significant (log-rank $p < 0.001$). Median survival decreased substantially from early to late stages, consistent with established lung cancer progression patterns.

Table 1. Baseline characteristics of patients with lung cancer (N = 13,402)

Characteristic	Value
Sample size	N = 13,402
Age (years)	Mean = 64.09, SD = 9.50, Range = 18–94
Age groups (1–4)	Mean = 2.94 (SD = 0.86)
Gender	Male: 79.4% (10,642) — Female: 20.6% (2,760)
Nationality	Mean = 1.65 (SD = 0.75) *(3 categories)*
Place of residence	Urban: 62.2% (8,334) — Rural: 37.8% (5,068)
Social status (1–3)	Mean = 1.92, SD = 0.75
Circumstances of diagnosis (1–3)	Mean = 1.74, SD = 0.60
Histology	Squamous carcinoma (most common): 34.9% (4,418)
Stage	Stage III (most common): 47.5% (6,370)
Overall survival time (days)	Median = 167, IQR = 53–399
Death (overall survival event)	75.9% (10,166)
Cancer-specific death (CSS)	46.2% (6,197)
Survival time range	1 to 1823 days

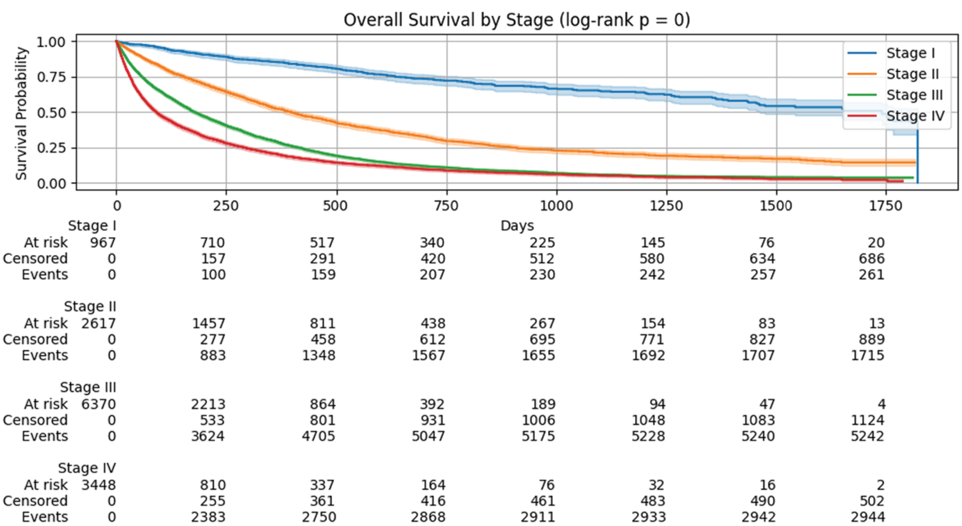


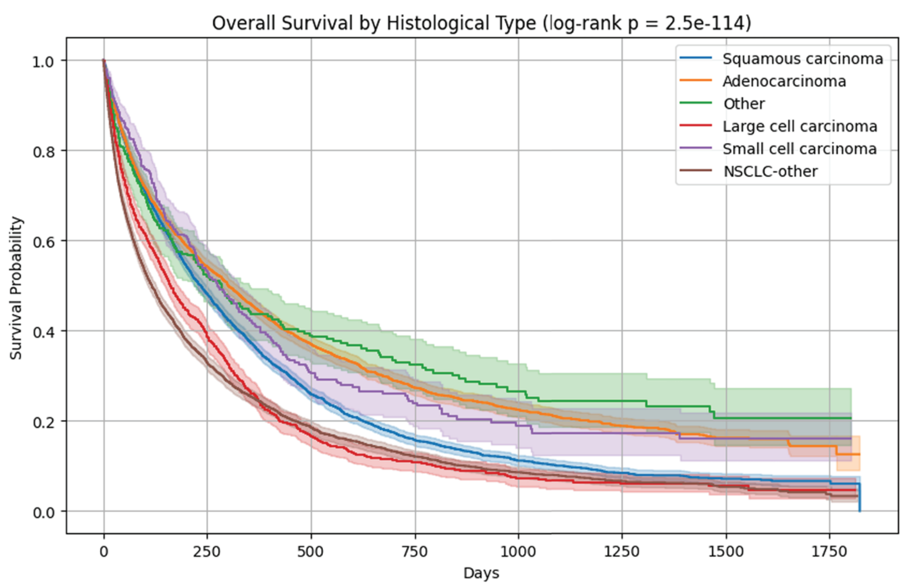
Figure 2. Kaplan–Meier survival curves by tumor stage (I–IV). Kaplan–Meier curves showing overall survival stratified by tumor stage (I–IV). A progressive decline in survival is observed with advancing stage. The difference in survival among the stages was statistically significant (log-rank $p < 0.001$). Number-at-risk tables are shown below the plot.

Survival by histological type

Overall survival varied significantly across histological subtypes (log-rank $p < 0.001$) (Figure 3). Patients with adenocarcinoma showed comparatively better survival trends, whereas those with squamous cell carcinoma, large cell carcinoma, NSCLC otherwise specified, and other categories exhibited steeper declines in survival probability over time. These findings highlight the prognostic relevance of tumor histopathology in lung cancer survival.

Survival by place of residence

Patients residing in urban areas demonstrated significantly better survival compared with those living in rural regions (log-rank $p < 0.001$) (Figure 4). This



		0	250	500	750	1000	1250	1500	1750
Squamous carcinoma	At risk	4418	1858	832	395	219	114	65	11
	Censored	0	380	613	757	831	891	925	976
	Events	0	2180	2973	3266	3368	3413	3428	3431
Adenocarcinoma	At risk	3693	1672	922	516	313	184	87	10
	Censored	0	421	690	885	1007	1093	1170	1242
	Events	0	1600	2081	2292	2373	2416	2436	2441
Other	At risk	311	141	88	56	36	24	15	3
	Censored	0	27	47	66	77	86	92	104
	Events	0	143	176	189	198	201	204	204
Large cell carcinoma	At risk	1247	395	138	73	31	18	8	2
	Censored	0	132	184	204	227	236	245	250
	Events	0	720	925	970	989	993	994	995
Small cell carcinoma	At risk	347	149	74	45	24	16	11	4
	Censored	0	48	64	78	91	97	101	108
	Events	0	150	209	224	232	234	235	235
NSCLC-other	At risk	3386	975	475	249	134	69	36	9
	Censored	0	214	313	389	441	479	501	521
	Events	0	2197	2598	2748	2811	2838	2849	2856

Figure 3. Kaplan–Meier survival curves by histological type. Kaplan–Meier curves comparing overall survival among histological subtypes of lung cancer. Significant heterogeneity in survival patterns is observed across adenocarcinoma, squamous cell carcinoma, small cell carcinoma, large cell carcinoma, NSCLC otherwise specified, and other histologies (log-rank $p < 0.001$). Number-at-risk tables are included.

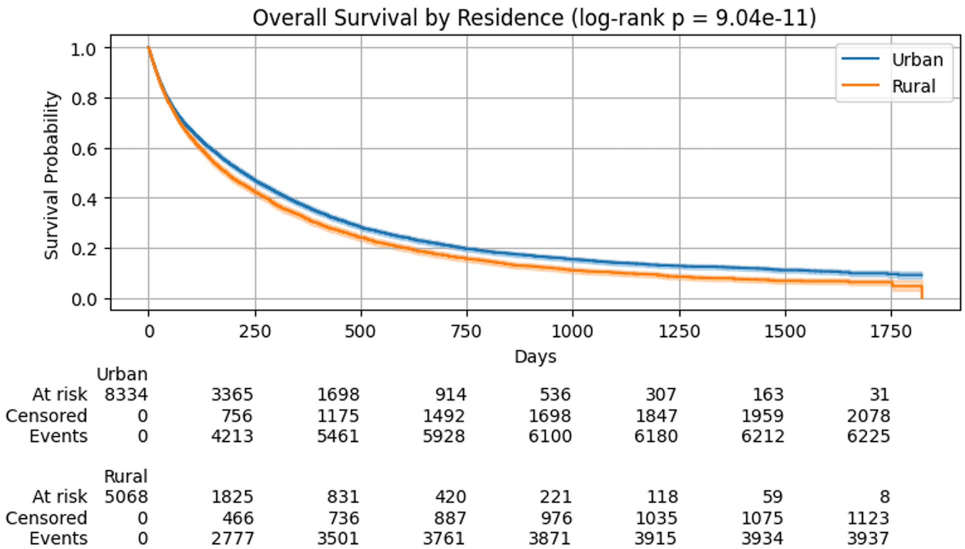


Figure 4. Kaplan–Meier survival curves by place of residence. Survival curves comparing overall survival between urban and rural residents. Urban patients experienced significantly improved survival compared with rural patients (log-rank $p < 0.001$). Number-at-risk tables appear beneath the curves.

disparity may reflect differences in healthcare access, diagnostic timing, and socioeconomic conditions.

Univariate Cox regression

In the univariate Cox proportional hazards analysis, several demographic, clinical, and tumor-related variables were significantly associated with overall survival (Table 2). Age was positively associated with mortality, with each additional year of age increasing the hazard of death. Male sex showed significantly worse survival compared with females. Patients residing in rural areas exhibited higher mortality than those living in urban regions. Lower social status and unfavorable diagnostic circumstances were also associated with increased risk of death. Histological subtype demonstrated substantial prognostic variation. Compared with adenocarcinoma, squamous cell carcinoma, large cell carcinoma, and other NSCLC subtypes were associated with significantly higher mortality. Tumor stage showed the strongest univariate effect, with hazard ratios increasing sharply from Stage I to Stage IV, reflecting progressively poorer survival with advancing disease.

Multivariable Cox regression

Independent prognostic factors

In the multivariable Cox proportional hazards model, several demographic, socioeconomic, and tumor-related characteristics remained independently associated with overall survival after adjustment for all covariates. Age was a significant continuous predictor, with increasing age associated with higher mortality. Gender remained a strong prognostic factor; male patients exhibited substantially higher mortality than females. Place of residence demonstrated a notable effect: urban residency was associated with a significantly reduced risk of death compared with rural residency. Social status showed a clear socioeconomic gradient, with lower social status categories (higher ordinal codes) associated with increased hazard of death. Histological subtype played an important prognostic role. Compared with adenocarcinoma (reference), certain subtypes—particularly large cell carcinoma, NSCLC otherwise specified, and squamous cell carcinoma—were associated with higher mortality risk. Tumor stage was the strongest independent predictor. Relative to Stage I, the hazard of death

Table 2. Univariate Cox proportional hazards regression analysis of factors associated with overall survival

Variable	Category / Comparison	HR	95% CI	p-value
Age	Per 1-year increase	1.02	1.01–1.02	<0.001
Age group (1–4)	Per increase in age group	1.17	1.14–1.19	<0.001
Circumstances of diagnosis	Per category increase (1–3)	1.19	1.15–1.23	<0.001
Gender	Male vs Female	1.56	1.48–1.64	<0.001
Histology	Large cell carcinoma vs Adenocarcinoma	1.66	1.54–1.79	<0.001
	NSCLC otherwise specified vs Adenocarcinoma	1.79	1.69–1.88	<0.001
	Other vs Adenocarcinoma	0.95	0.82–1.09	0.470
	Small cell carcinoma vs Adenocarcinoma	1.05	0.92–1.20	0.500
	Squamous carcinoma vs Adenocarcinoma	1.29	1.22–1.35	<0.001
Nationality	Per category increase	1.04	1.01–1.06	0.004
Place of residence	Urban vs Rural	0.88	0.84–0.91	<0.001
Social status (1–3)	Per increase in social status category	1.05	1.02–1.07	<0.001
Tumor stage	Stage II vs Stage I	3.35	2.94–3.82	<0.001
	Stage III vs Stage I	6.36	5.62–7.21	<0.001
	Stage IV vs Stage I	8.47	7.46–9.62	<0.001

Abbreviations: HR = hazard ratio; CI = confidence interval; NSCLC = non-small cell lung cancer.

increased sharply for Stages II, III, and IV, consistent with disease progression severity.

Adjusted hazard ratios and confidence intervals

Adjusted hazard ratios (HRs), 95% confidence intervals (CIs), and p-values from the multivariable Cox model are presented in Table 3. These results highlight the independent contribution of each covariate to overall survival while accounting for potential confounding across demographic, clinical, and pathological characteristics.

Model diagnostics

We conducted a comprehensive diagnostic evaluation to assess the validity of the Cox proportional hazards model. The proportional hazards assumption was evaluated using scaled Schoenfeld residuals and global tests (Figure 5). Several variables—including diagnostic circumstances, selected histological subtypes, and advanced tumor stages—showed statistical evidence of PH violation. However, graphical inspection revealed only minor deviations, and given the large sample size, these violations were considered acceptable. The standard Cox model was retained.

Martingale residual plots were used to examine the linearity of age as a continuous covariate (Figure 6A). The scatterplot showed no major systematic deviation from linearity, supporting the inclusion of age as a linear term in the final Cox model. Variance inflation factor (VIF) values for all covariates were below thresholds indicating problematic multicollinearity (Table 4). No evidence of strong linear dependency was detected among predictors. All variance inflation factor (VIF) values were below the commonly accepted threshold of 5, indicating no evidence of problematic multicollinearity among predictors in the multivariable Cox proportional hazards model. Delta-beta ($\Delta\beta$) residuals were examined to identify influential observations with disproportionate impact on model coefficients (Figure 6B). No individual data points exhibited undue influence on the overall model, indicating stability of the regression estimates.

Table 3. Multivariable Cox proportional hazards regression analysis of factors associated with overall survival

Variable	Category / Comparison	Adjusted HR	95% CI	p-value
Age	Per 1-year increase	1.02	1.01–1.02	<0.001
Gender	Male vs Female	1.49	1.41–1.57	<0.001
Nationality	Per category increase	1.02	1.00–1.05	0.070
Place of residence	Urban vs Rural	0.87	0.84–0.91	<0.001
Social status (1–3)	Per one-category decrease (from high to low)	1.07	1.05–1.10	<0.001
Circumstances of diagnosis	Per category increase	1.14	1.10–1.18	<0.001
Histology	Large cell carcinoma vs Adenocarcinoma	1.33	1.23–1.43	<0.001
Histology	NSCLC otherwise specified vs Adenocarcinoma	1.50	1.42–1.58	<0.001
Histology	Other vs Adenocarcinoma	1.10	0.95–1.27	0.190
Histology	Small cell carcinoma vs Adenocarcinoma	1.03	0.90–1.17	0.719
Histology	Squamous carcinoma vs Adenocarcinoma	1.10	1.04–1.16	<0.001
Tumor stage	Stage II vs Stage I	2.92	2.56–3.33	<0.001
Tumor stage	Stage III vs Stage I	5.37	4.73–6.09	<0.001
Tumor stage	Stage IV vs Stage I	7.70	6.78–8.75	<0.001

Discussion

In this large, population-based cohort of 13,402 lung cancer patients from the National Cancer Registry of Kazakhstan, we found that overall survival was poor, with a median OS of 167 days and nearly three-quarters of patients dying during follow-up. Older age, male sex, rural residence, lower social status, unfavorable diagnostic circumstances, non-adenocarcinoma histology, and advanced stage at diagnosis were all associated with higher mortality in univariate analyses. In multivariable Cox models, age, male sex, rural residence, lower social status, diagnostic circumstances, histology, and stage remained independent prognostic factors. Among these, tumor stage and histological subtype were the strongest predictors, while residence and social status highlighted important socio-geographic disparities in survival. Our findings are consistent with global evidence that lung cancer remains one of the leading causes of cancer mortality and that prognosis is strongly determined by stage and histology at diagnosis. Large contemporary cohort studies from Europe and Asia likewise report that advanced TNM stage is the dominant determinant of survival, even in settings with increased access to surgery and systemic therapies (7). The shorter survival we observed, particularly for stage III–IV disease, is in line with data from other middle-income settings, where late presentation and limited access to curative treatment are common (8). The prognostic role of histological subtype in our cohort—worse survival for large cell, NSCLC otherwise specified, and squamous carcinoma compared with adenocarcinoma—also mirrors recent reports. A 2021 SEER-based analysis demonstrated distinct survival trajectories across histological types, with adenocarcinoma generally showing better outcomes than squamous and other non-small-cell variants (9). More recent work has emphasized that histological heterogeneity interacts with molecular features and treatment modalities, further widening survival differences between subgroups (10). Age and sex effects in our study—higher mortality with older age and among men—are compatible with international data. Population-based series and screening cohorts consistently show that older age at diagnosis is associated with poorer OS, while male sex remains a negative prognostic factor even after adjustment for smoking and comorbidities (11). In Kazakhstan, national statistics have also reported a predominance of lung cancer among men and in older age groups,

ACTA BIOMED 2026; VOL. 97, N. 1: 18341 DOI: 10.23750/ABM.2026.18341

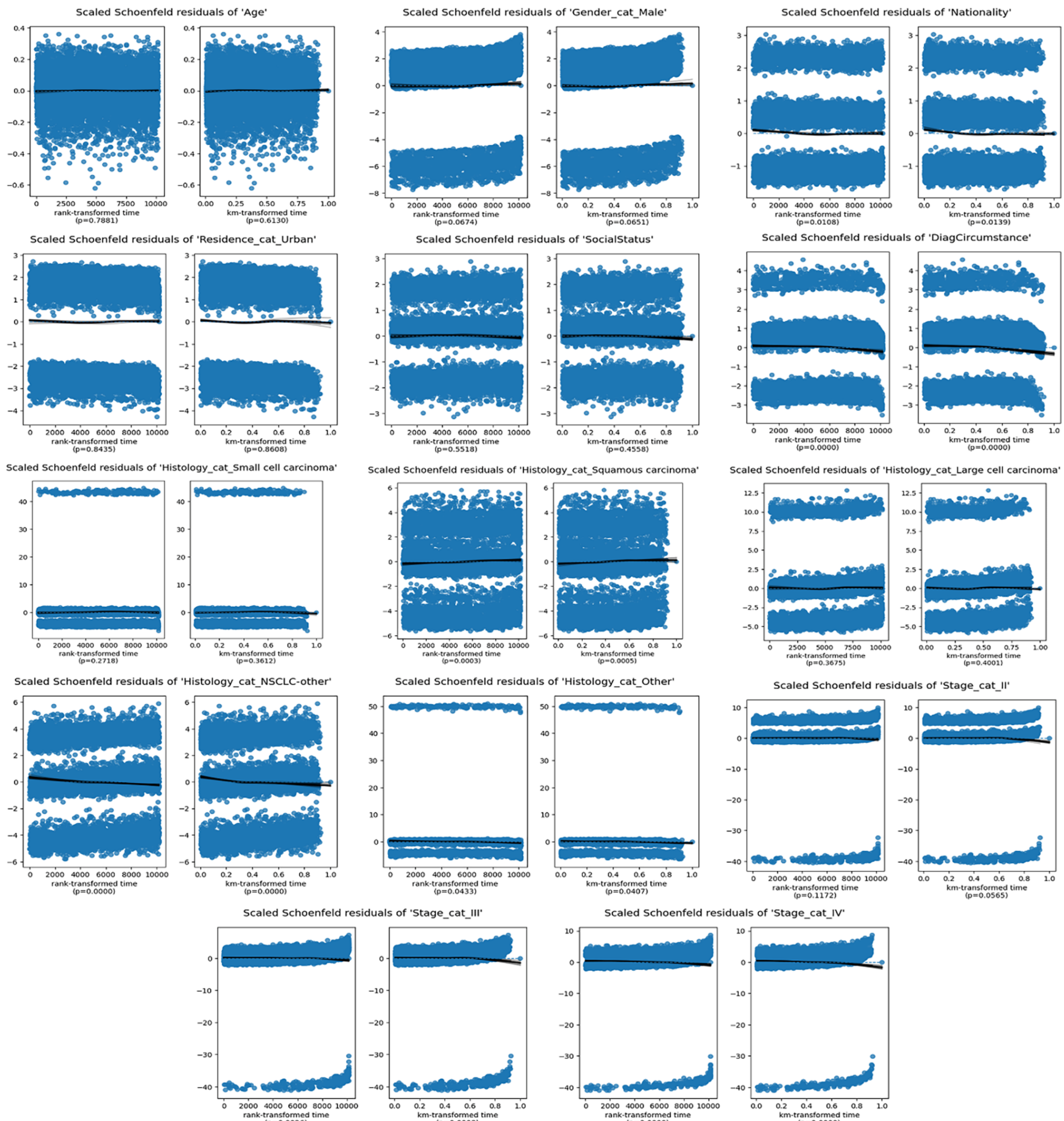


Figure 5. Diagnostic plots for the Cox proportional hazards model. Scaled Schoenfeld residuals for selected covariates. Residual patterns show minor deviations in a subset of predictors; however, no clinically meaningful violation of the proportional hazards assumption was identified.

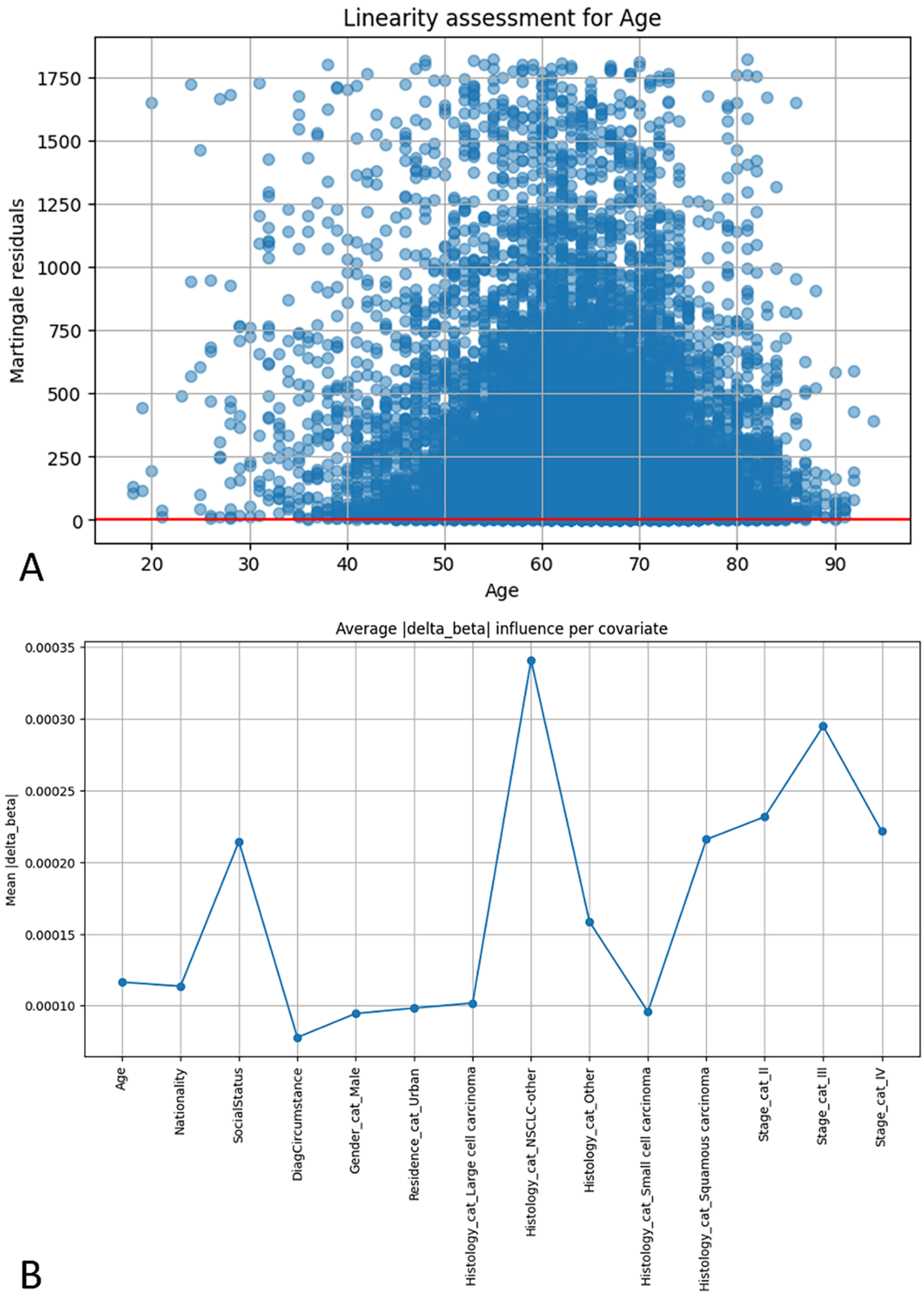


Figure 6. Diagnostic plots for the Cox proportional hazards model. A) Martingale residual plot assessing the linearity of age. No systematic nonlinear trend was observed, supporting the use of age as a linear continuous predictor. B) Mean absolute delta-beta (Δβ) residuals across covariates. No covariate exhibited unusually high Δβ values, indicating the absence of influential observations and confirming the stability of the multivariable Cox model.

Table 4. Variance inflation factor (VIF) values for covariates in the multivariable Cox model

Variable	VIF
Age	1.05
Gender (Male)	1.08
Residence (Urban)	1.04
Social status	1.05
Diagnostic circumstances	1.04
Histology (dummy set)	1.60
Stage (dummy set)	4.23

which our study confirms at the survival level (3). The marked urban–rural survival gap we observed is particularly important in the Kazakhstani context. Recent work from high-income countries has shown that rural residents often experience higher lung cancer mortality, slower declines in death rates, and less access to modern diagnostics and therapies (12). A 2025 analysis of Black and White populations in the United States found persistent rural–urban disparities in lung cancer incidence and mortality, with rural areas bearing a disproportionate burden (12). Beyond treatment access, environmental exposures such as indoor biomass smoke and solid fuel use, which are more prevalent in rural settings, have been linked to excess lung cancer risk and may also contribute to poorer outcomes (13). In Kazakhstan, recent studies have highlighted broader inequalities in healthcare accessibility between urban and rural populations, supporting our interpretation that survival differences may partly reflect structural barriers to timely diagnosis and optimal care (14). Our findings regarding social status align with a growing body of evidence that socioeconomic position influences both lung cancer incidence and survival. Large multi-country cohort analyses and umbrella reviews have shown that lower socioeconomic status is associated with higher lung cancer incidence and worse survival, even after accounting for smoking and stage (15). Recent work from East Asia and Europe has confirmed that income, education, and area-based deprivation indices predict poorer outcomes across several cancers, including lung cancer (16). These data support our interpretation that the gradient in hazard across social status categories reflects underlying inequities in risk exposure, health literacy, and access to guideline-concordant treatment. Within Kazakhstan, previous epidemiological work has described rising cancer burden and geographic heterogeneity in lung cancer incidence and mortality, but has provided limited survival detail (3). Our study adds to this literature by quantifying how demographic, clinical, and socioeconomic variables jointly shape survival in a recent national cohort. While earlier Kazakhstani research has focused on molecular prognostic markers in selected NSCLC populations (17), and on early detection strategies such as low-dose CT screening (18), our registry-based analysis provides complementary system-level insight into real-world outcomes across all histological types and stages.

Clinical and public health implications

The very short median OS in our cohort and the strong prognostic impact of stage at diagnosis underscore the urgent need to improve early detection of lung cancer in Kazakhstan. International studies show that organized screening with low-dose CT in high-risk populations can shift stage distribution towards earlier disease and improve survival (18). However, implementation in low- and middle-income settings is challenged by resource constraints, competing health priorities, and workforce capacity (5, 12). Our results suggest that any future Kazakhstani screening initiatives should be carefully targeted and accompanied by efforts to reduce diagnostic delays, particularly in rural and socioeconomically disadvantaged populations. The observed disparities by residence and social status highlight the need for a stronger equity focus in cancer control strategies. Recent analyses from other countries show that interventions improving access to early diagnosis, multidisciplinary care, and financial protection can mitigate SES-related survival gaps (19). In Kazakhstan, this could translate into expanding oncology services outside major urban centers, strengthening referral pathways from primary care, and ensuring that rural and low-income patients have timely access to imaging, pathology, and systemic therapies. The strong and independent prognostic effect of histological subtype also has therapeutic implications. As precision oncology advances, histology remains a key anchor for decisions around surgery, radiotherapy, and systemic treatment, and is increasingly complemented by molecular profiling (10). Ensuring high-quality pathological classification and access to essential biomarker testing in Kazakhstan will be critical to align outcomes with international standards.

Strengths and limitations

This study has several strengths. First, it is based on a large, population-level cohort of over 13,000 patients drawn from a national cancer registry, enhancing generalizability within Kazakhstan. Second, we used rigorous survival methods, including Kaplan–Meier estimation and multivariable Cox regression, supported by comprehensive model diagnostics (PH assumption, martingale residuals, VIF, and delta-beta). Third, we incorporated both clinical and socio-demographic variables, allowing us to highlight equity-relevant determinants such as residence and social status, which are often under-reported in survival analyses from the region. Nevertheless, important limitations should be acknowledged. The retrospective design inherently limits control over data quality and confounding. We lacked information on treatment modalities (surgery, radiotherapy, systemic therapies), comorbidities, smoking status, and performance status, all of which are known to influence lung cancer outcomes and may partially mediate the associations we observed (20). Residual confounding by unmeasured factors such as environmental exposures is also likely (15). In addition, the social status and nationality variables were relatively crude proxies for socioeconomic position and ethnicity, which may underestimate the true magnitude of socioeconomic gradients. Finally, while we conducted extensive PH diagnostics, some covariates showed statistical violations; although we judged these deviations to be minor and acceptable in light of the sample size, they may still have introduced small biases in HR estimates. The national cancer registry dataset lacked information on smoking history, comorbidities, performance status, and treatment modalities, which are known determinants of lung cancer prognosis and could not be evaluated in the present analysis.

Future research directions

Future work should integrate clinical and treatment data from hospital information systems to disentangle the contributions of therapy, comorbidity, and health system factors to survival differences in Kazakhstan. Prospective or quasi-prospective cohorts could provide more granular information on smoking history, occupational and environmental exposures, and comorbidity burden, thereby refining risk stratification (20). There is also a strong rationale for developing and validating multivariable prediction models and machine-learning–based tools for survival estimation and clinical decision support, building upon recent methodological work in Kazakhstan and elsewhere (14). Finally, implementation research is needed to evaluate how targeted interventions—such as low-dose CT screening pilots, tele-oncology support for rural regions, or financial protection schemes for disadvantaged patients—affect stage at diagnosis and survival disparities over time (12).

Conclusion

In this nationwide cohort of 13,402 lung cancer patients in Kazakhstan, we observed poor overall survival and identified age, sex, place of residence, social status, diagnostic circumstances, histological subtype, and tumor stage as independent determinants of mortality. Advanced stage and non-adenocarcinoma histologies were associated with the highest risks, while rural residence and lower social status highlighted significant equity gaps. These findings underscore the urgent need to improve early detection, strengthen access to high-quality cancer care—particularly for rural and socioeconomically disadvantaged groups—and incorporate histological and sociodemographic information into risk-stratified management strategies. Integrating these insights into Kazakhstan's national cancer control policies and future screening and treatment initiatives has the potential to meaningfully reduce lung cancer mortality and narrow survival disparities.

Acknowledgments: The authors gratefully acknowledge the Republican Center for Healthcare Development under the Ministry of Health of the Republic of Kazakhstan, in the person of Deputy Chairman of the Board Mr. S.S. Shaikhiyev, for the provision of essential data used in this study.

Funding: This research received no external funding.

Ethic approval: The study was conducted in accordance with the Declaration of Helsinki and approved by the Bioethics Committee of West Kazakhstan Medical University (Protocol No. 3, March 28, 2025).

Informed consent statement: Patient consent was waived due to the use of anonymized registry data.

Data availability statement: The data presented in this study are available from the corresponding authors upon reasonable request. Restrictions apply due to privacy and ethical considerations.

Conflicts of interest: Each author declares that he or she has no commercial associations (e.g., consultancies, stock ownership, equity interest, patent/licensing arrangement, etc.) that might pose a conflict of interest in connection with the submitted article.

Author contributions: Conceptualization, Z.A., A.M., Y.I., N.M. and A.T.; methodology, Z.A.; resources, A.M.; data curation, Z.A.; formal analysis, Z.A.; investigation, Z.A.; writing—original draft preparation, Z.A. and A.T.; writing—review and editing, Z.A., A.M., Y.I., N.M., and A.T.; supervision, A.M., Y.I. All authors have read and agreed to the published version of the manuscript.

Declaration on the Use of AI: None.

References

1. Thandra KC, Barsouk A, Saginala K, Aluru JS, Barsouk A. Epidemiology of lung cancer. Contemp Oncol. 2021;25(1):45-52. doi:10.5114/wo.2021.103829
2. Cao Z, He L, Luo Y, Tong X, et al.. Burden of chronic respiratory diseases and their attributable risk factors in 204 countries and territories, 1990-2021: Results from the global burden of disease study 2021. Chin Med J Pulm Crit Care Med. 2025;3(2):100-10. doi:10.1016/j.pccm.2025.05.005
3. Yessenbayev D, Khamidullina Z, Tarzhanova D, et al. Epidemiology of lung cancer in Kazakhstan: trends and geographic distribution. Asian Pac J Cancer Prev. 2023;24(5):1521-32. doi:10.31557/APJCP.2023.24.5.1521
4. Ramos R, Moura CS, Costa M, et al. Heterogeneity of lung cancer: The histopathological diversity and tumour classification in the Artificial Intelligence era. Pathobiology. 2025;92(4):239-50. doi:10.1159/000544892
5. Lubuzo B, Ginindza T, Hlongwana K. The barriers to initiating lung cancer care in low-and middle-income countries. Pan Afr Med J. 2020;35:38. doi:10.11604/pamj.2020.35.38.17333
6. Spankulova L, Chulanova Z, Konyrbay A. Evaluating healthcare accessibility in Kazakhstan: urban and rural perspectives. Eurasian Journal of Economic and Business Studies. 2024;68(2):5-19. doi:10.47703/ejebs.v68i2.376
7. Teixidor-Vila E, Trallero J, Puigdemont M, et al. Lung cancer survival trends and prognostic factors: A 26-year population-based study in Girona Province, Spain. Lung Cancer. 2024;197:107995. doi:10.1016/j.lungcan.2024.107995
8. Torrente M, Sousa PA, Franco F, et al. Understanding prognosis and survival outcomes in patients with early-stage non-small-cell lung cancer. Clin Med. 2022;22(Suppl 4): 38-40. doi:10.7861/clinmed.22-4-s38
9. Hao B, Fan T, Xiong J, et al. The prognostic significance of the histological types in patients with nonsmall cell lung cancer≤ 2 cm. Front Surg. 2021;8:721567. doi:10.3389/fsurg.2021.721567
10. Arici MO, Guzel HG, Salim DK, et al. Prognostic factors and survival data of stage I lung cancer: A single-center experience. J Clin Pract Res. 2025;47(5):519-27. doi:10.14744/cpr.2025.58353
11. Li J, Xu HL, Li WX, Ma XY, Liu XH, Zhang ZF. Prognostic factors of survival in patients with lung cancer after low-dose computed tomography screening: a multivariate analysis of a lung cancer screening cohort in China. BMC Cancer. 2025;25(1):646. doi:10.1186/s12885-025-14036-9
12. Howlader N, Cronin KA, Yu M, Miller D, Lowy DR. Urban-rural disparities in lung cancer incidence and mortality patterns in Black and White populations. Cancer. 2025;131(15):e70004. doi:10.1002/cncr.70004
13. Yang R, Mu X, Wang G, et al. Urban-rural inequality in lung cancer risk from indoor air pollution: Prolonged indoor time amplifies the risks of solid fuel use. J Environ Manage. 2025;393:127084. doi:10.1016/j.jenvman.2025.127084
14. Kar I, Vhora F, Bou Zerdan M, et al. Survival determinants and sociodemographic disparities in early-onset non–small cell lung cancer. JAMA Netw Open. 2025;8(10):e2537307. doi:10.1001/jamanetworkopen.2025.37307
15. Onwuka JU, Zahed H, Feng X, et al. Association between socioeconomic position and lung cancer incidence in 16 countries: a prospective cohort consortium study. EClinicalMedicine. 2025;82:103152. doi:10.1016/j.eclinm.2025.103152
16. Lee J, Park J, Kim N, et al. Socioeconomic disparities in six common cancer survival rates in South Korea: population-wide retrospective cohort study. JMIR Public Health Surveill. 2024;10:e55011. doi:10.2196/55011
17. Yessentayeva SY, Makarov VA, Kalmatayeva ZA, Zhakenova ZK, Arybzhanov DT. Molecular genetic tests in survival

factors in patients with NSCLC in the clinical practice of Kazakhstan. Med J Islam Repub Iran. 2021;35:133. doi:10.47176/mjiri.35.133

18. Panina A, Kaidarova D, Zholdybay Z, et al. Lung cancer screening with low-dose chest computed tomography: experience from radon-contaminated regions in Kazakhstan. J Prev Med Public Health. 2022;55(3):273-9. doi:10.3961/jpmph.21.600
19. Li S, He Y, Liu J, et al. An umbrella review of socioeconomic status and cancer. Nat Commun. 2024;15(1):9993. doi:10.1038/s41467-024-54444-2
20. Buja A, Di Pumpo M, Rugge M, et al. Patterns of comorbidities in lung cancer patients and survival. Cancers. 2025;17(9):1577. doi:10.3390/cancers17091577