

Supplementary

Appendix 1 Supplementary Methods

This population-based retrospective cohort study was conducted using the SEERStat Database: Incidence – SEER Research Data, 17 Registries, Nov 2024 Sub (2000–2022). Anatomical segmentectomy was identified using the SEER variable “RX_SUMM_SURG_PRIM_SITE”, coded as B220 (stored as 220 in SEER*Stat exports). Only patients with single primary NSCLC who underwent segmentectomy were included.

Tumor size was extracted from “Tumor Size Over Time Recode (1988+) (mm)”. Because detailed AJCC 8th edition T1a/b/c subclassification was not consistently available across all years, tumor maximum diameter was used to reconstruct size-based T categories. Tumors measuring 1–10 mm, 11–20 mm, and 21–30 mm were classified as T1a, T1b, and T1c, respectively, and only tumors ≤ 30 mm were retained. Among tumors measuring ≤ 3.0 cm, some cases were reconstructed as T2–4 under the 8th edition of the American Joint Committee on Cancer staging system (The AJCC 8th edition). These cases were excluded because T categorization in the AJCC system is based not only on tumor size, but also on other invasion- or extension-related descriptors. Therefore, eligibility for the final cohort was determined by reconstructed AJCC 8th edition T1N0M0 status rather than tumor size alone. Clinical nodal information in SEER was used to restrict the cohort to N0M0 disease.

Regional lymph node yield was defined using the SEER variable “Regional nodes examined (1988+),” representing the number of regional lymph nodes pathologically evaluated. Cases with Regional nodes examined coded as 95–98 were excluded because the exact number of examined lymph nodes could not be reliably quantified. This variable reflects the extent of nodal assessment but does not distinguish systematic dissection from sampling. Lymph node yield was analyzed both categorically (0, 1–5, 6–10, and >10) and continuously.

Overall survival time was derived from the SEER variable “survival months,” and survival status was determined according to “Vital status recode.” For competing risk analyses, cause of death was defined using “SEER Cause-specific death classification” and “SEER Other cause of death classification,” coded as 0 = alive, 1 = death due to lung cancer, and 2 = death due to other causes. Lung cancer-specific death was treated as the event of interest, with other-cause death as the competing event.

Multivariable models adjusted for age, sex, race, tumor size, tumor grade, histologic subtype, radiotherapy, chemotherapy, laterality, and primary tumor lobe. Age and tumor size were

modeled as continuous variables; in interaction analyses, age was stratified at the cohort median (70 years), and tumor size was stratified into 1–10, 11–20, and 21–30 mm categories. Overall survival was analyzed using Kaplan–Meier methods and Cox proportional hazards models. Dose–response relationships were evaluated using restricted cubic spline models with four knots placed according to Harrell’s recommended percentile-based approach. Effect modification was assessed using interaction terms within the Cox framework. Competing risk analyses were performed using Fine–Gray sub-distribution hazard models, and cumulative incidence functions were compared using Gray’s test.

Analyses were conducted using R Statistical Software (Version 4.2.2, <http://www.R-project.org>, The R Foundation) and the Free Statistics analysis platform (Version 2.3, Beijing, China). All tests were two-sided, with $P < 0.05$ considered statistically significant.

Table S1 Full multivariable Cox model (all covariates)

Variable	Adjusted HR (95% CI)	P value
Number of examined lymph nodes (LNY)		
(reference: Group 1)		
Group 2	0.79 (0.67–0.94)	0.006
Group 3	0.55 (0.44–0.70)	<0.001
Group 4	0.40 (0.30–0.52)	<0.001
P for trend	0.74 (0.68–0.8)	<0.001
Age (per year increase)	1.04 (1.03–1.05)	<0.001
Sex (reference: male)	0.77 (0.67–0.89)	<0.001
Race (reference: White)		
Black	0.90 (0.67–1.20)	0.465
Other/unknown	0.63 (0.45–0.89)	0.008
Tumor size (reference: 1–10 mm)		
11–20 mm	1.34 (1.06–1.69)	0.015
21–30 mm	1.85 (1.44–2.38)	<0.001
Laterality (reference: left)		
Right	1.06 (0.92–1.23)	0.397
Other	0 (0–Inf)	0.987
Primary tumor lobe (reference: upper lobe)		

Middle lobe	1.09 (0.72–1.64)	0.69
Lower lobe	0.96 (0.82–1.11)	0.554
Unknown	0.68 (0.32–1.46)	0.326
Tumor grade (reference: I–II)		
III–IV	1.35 (1.14–1.61)	<0.001
Unknown	0.89 (0.68–1.16)	0.39
Histology (reference: adenocarcinoma)		
Squamous cell carcinoma	1.71 (1.46–2.02)	<0.001
Other NSCLC	1.30 (0.95–1.77)	0.096
Adjuvant treatment		
Radiotherapy (yes vs. no)	1.93 (1.37–2.71)	<0.001
Chemotherapy (yes vs. no)	2.07 (1.31–3.28)	0.002

Hazard ratios (HR) and 95% confidence intervals (CI) were estimated using multivariable Cox proportional hazards regression models. Group 1 represents patients with no examined lymph nodes (0 nodes); Group 2, 1–5 nodes; Group 3, 6–10 nodes; and Group 4, more than 10 nodes examined. The multivariable model was adjusted for age, sex, race, tumor size, tumor grade, histologic subtype, radiotherapy, chemotherapy, laterality, and primary tumor lobe. P for trend was calculated by modeling the lymph node group as an ordinal variable in the Cox regression model. HR, hazard ratio; CI, confidence interval; NSCLC, non–small cell lung cancer.

Table S2 Proportional hazards assumption tests

Model	Test	χ^2	df	P value	Interpretation
Multivariable Cox (primary model)	Schoenfeld residuals (global)	5.269	3	0.153	No evidence of PH violation*

PH, proportional hazards. Proportional hazards assumption was assessed using global Schoenfeld residual tests.

Table S3 Multivariable Cox model with additional adjustment for diagnosis year

Variable	Adjusted HR (95% CI)	P value
Regional lymph node yield (reference: Group 1)		
Group 2	0.8 (0.68–0.95)	0.012

Group 3	0.6 (0.47–0.76)	<0.001
Group 4	0.44 (0.33–0.58)	<0.001
P for trend	0.76 (0.7–0.83)	<0.001
Age (per year increase)	1.04 (1.03–1.05)	<0.001
Sex (reference: male)	0.78 (0.67–0.9)	0.001
Race (reference: White)		
Black	0.89 (0.67–1.19)	0.443
Other/unknown	0.66 (0.47–0.93)	0.018
Year of diagnosis (2011–2022 vs. 2000–2010)	0.74 (0.63–0.86)	<0.001
Tumor size (reference: 1–10 mm)		
11–20 mm	1.29 (1.02–1.64)	0.031
21–30 mm	1.76 (1.37–2.26)	<0.001
Laterality (reference: left)		
Right	1.06 (0.92–1.23)	0.412
Other	0 (0–Inf)	0.987
Primary tumor lobe (reference: upper lobe)		
Middle lobe	1.05 (0.7–1.58)	0.808
Lower lobe	0.97 (0.84–1.13)	0.719
Unknown	0.66 (0.31–1.41)	0.283
Tumor grade (reference: I–II)		
III–IV	1.32 (1.11–1.57)	0.001
Unknown	0.89 (0.68–1.16)	0.387
Histology (reference: adenocarcinoma)		
Squamous cell carcinoma	1.7 (1.44–2)	<0.001
Other NSCLC	1.28 (0.94–1.74)	0.12
Adjuvant treatment		
Radiotherapy (yes vs. no)	1.87 (1.33–2.64)	<0.001
Chemotherapy (yes vs. no)	2 (1.26–3.18)	0.003
Model adjusted for the same covariates as the primary analysis, with additional adjustment for year of diagnosis.		

Table S4 Multivariable Cox model in patients without radiotherapy or chemotherapy

Variable	Adjusted HR (95% CI)	P value
Regional lymph node yield (reference: Group 1)		
Group 2	0.78 (0.65–0.93)	0.005
Group 3	0.58 (0.45–0.74)	<0.001
Group 4	0.42 (0.31–0.55)	<0.001
P for trend	0.75 (0.69–0.82)	<0.001
Age (per year increase)	1.04 (1.03–1.05)	<0.001
Sex (reference: Male)	0.78 (0.67–0.9)	0.001
Race (reference: White)		
Black	0.99 (0.73–1.33)	0.926
Other/unknown	0.71 (0.51–1.01)	0.056
Year of diagnosis (2011–2022 vs. 2000–2010)	0.74 (0.63–0.87)	<0.001
Tumor size (reference: 1–10 mm)		
11–20 mm	1.33 (1.04–1.68)	0.021
21–30 mm	1.76 (1.36–2.27)	<0.001
Laterality (reference: left)		
Right	1.03 (0.89–1.2)	0.655
Other	0 (0–Inf)	0.987
Primary tumor lobe (reference: upper lobe)		
Middle lobe	1.01 (0.65–1.55)	0.981
Lower lobe	0.97 (0.83–1.13)	0.697
Unknown	0.65 (0.3–1.39)	0.264
Tumor grade (reference: I–II)		
III–IV	1.36 (1.14–1.62)	0.001
Unknown	0.9 (0.68–1.19)	0.446
Histology (reference: adenocarcinoma)		
Squamous cell carcinoma	1.71 (1.45–2.03)	<0.001

Other NSCLC

1.21 (0.87–1.68)

0.255

Model adjusted for the same covariates as the primary analysis, with additional adjustment for year of diagnosis; patients who received radiotherapy or chemotherapy were excluded.

Table S5 Multivariable Fine-Gray regression for lung cancer-specific death

Variable	Adjusted sdHR	Adjusted 95% CI	Adjusted P value
Age (per year)	1.01	1–1.02	0.132
Sex (female vs. male)	0.8	0.65–0.99	0.039
Race (reference: White)			
Black	0.73	0.48–1.1	0.13
Other/unknown	0.52	0.31–0.87	0.012
Tumor size (reference: ≤10 mm)			
11–20 mm	1.55	1.09–2.21	0.014
21–30 mm	2.14	1.46–3.14	<0.001
Tumor grade (reference: I–II)			
III–IV	1.57	1.23–2.01	<0.001
Unknown	1.03	0.71–1.47	0.892
Histology (reference: adenocarcinoma)			
Squamous cell carcinoma	0.89	0.69–1.16	0.4
Other NSCLC	1.26	0.8–1.99	0.32
Adjuvant treatment			
Radiotherapy (yes vs. no)	2.05	1.29–3.25	0.002
Chemotherapy (yes vs. no)	3.39	2.19–5.27	<0.001
Regional lymph node yield (LNY) (reference: Group 1)			
Group 2	0.74	0.57–0.94	0.016
Group 3	0.58	0.41–0.8	0.001
Group 4	0.43	0.29–0.63	<0.001

sdHR, sub-distribution hazard ratio. Multivariable Fine-Gray models were adjusted for age, sex, race, tumor size, tumor grade, histology, radiotherapy, chemotherapy, laterality, primary lobe, and LNY. LNY, regional lymph node yield.

Table S6 Multivariable Fine-Gray regression for other-cause death

Variable	Adjusted sdHR	Adjusted 95% CI	Adjusted P value
Age (per year)	1.05	1.03–1.06	<0.001
Sex (female vs. male)	0.86	0.71–1.06	0.153
Race (reference: White)			
Black	1.2	0.82–1.76	0.35
Other/unknown	0.87	0.56–1.34	0.517
Tumor size (reference: ≤10 mm)			
11–20 mm	1.11	0.82–1.51	0.484
21–30 mm	1.36	0.98–1.88	0.069
Tumor grade (reference: I–II)			
III–IV	1	0.78–1.27	0.99
Unknown	0.79	0.54–1.17	0.247
Histology (reference: adenocarcinoma)			
Squamous cell carcinoma	2.41	1.94–2.98	<0.001
Other NSCLC	1.32	0.88–1.99	0.178
Adjuvant treatment			
Radiotherapy (yes vs. no)	1.05	0.58–1.92	0.868
Chemotherapy (yes vs. no)	0.24	0.05–1.05	0.057
Regional lymph node yield (LNY) (reference: Group 1)			
Group 2	0.91	0.73–1.15	0.439
Group 3	0.56	0.4–0.78	0.001
Group 4	0.41	0.28–0.61	<0.001

sdHR, sub-distribution hazard ratio. Multivariable Fine-Gray models were adjusted for age, sex, race, tumor size, tumor grade, histology, radiotherapy, chemotherapy, laterality, primary lobe, and LNY. LNY, regional lymph node yield.

Table S7 Pairwise log-rank comparisons of overall survival among regional lymph node yield groups

Pairwise comparison	Group sample sizes	Total n	Unadjusted P value	Holm-adjusted P value
------------------------	-----------------------	---------	-----------------------	--------------------------

Group 1 vs. Group 2	351 vs. 793	1,144	6.971×10^{-5}	1.394×10^{-4}
Group 1 vs. Group 3	351 vs. 538	889	8.226×10^{-13}	4.113×10^{-12}
Group 1 vs. Group 4	351 vs. 535	886	1.057×10^{-20}	6.341×10^{-20}
Group 2 vs. Group 3	793 vs. 538	1,331	2.332×10^{-5}	6.996×10^{-5}
Group 2 vs. Group 4	793 vs. 535	1,328	5.009×10^{-11}	2.003×10^{-10}
Group 3 vs. Group 4	538 vs. 535	1,073	0.011636	0.011636

P values were derived from pairwise log-rank tests. Adjusted P values were calculated across all six pairwise comparisons using the Holm method. Group 1, 0 examined nodes; Group 2, 1–5 nodes; Group 3, 6–10 nodes; Group 4, >10 nodes.

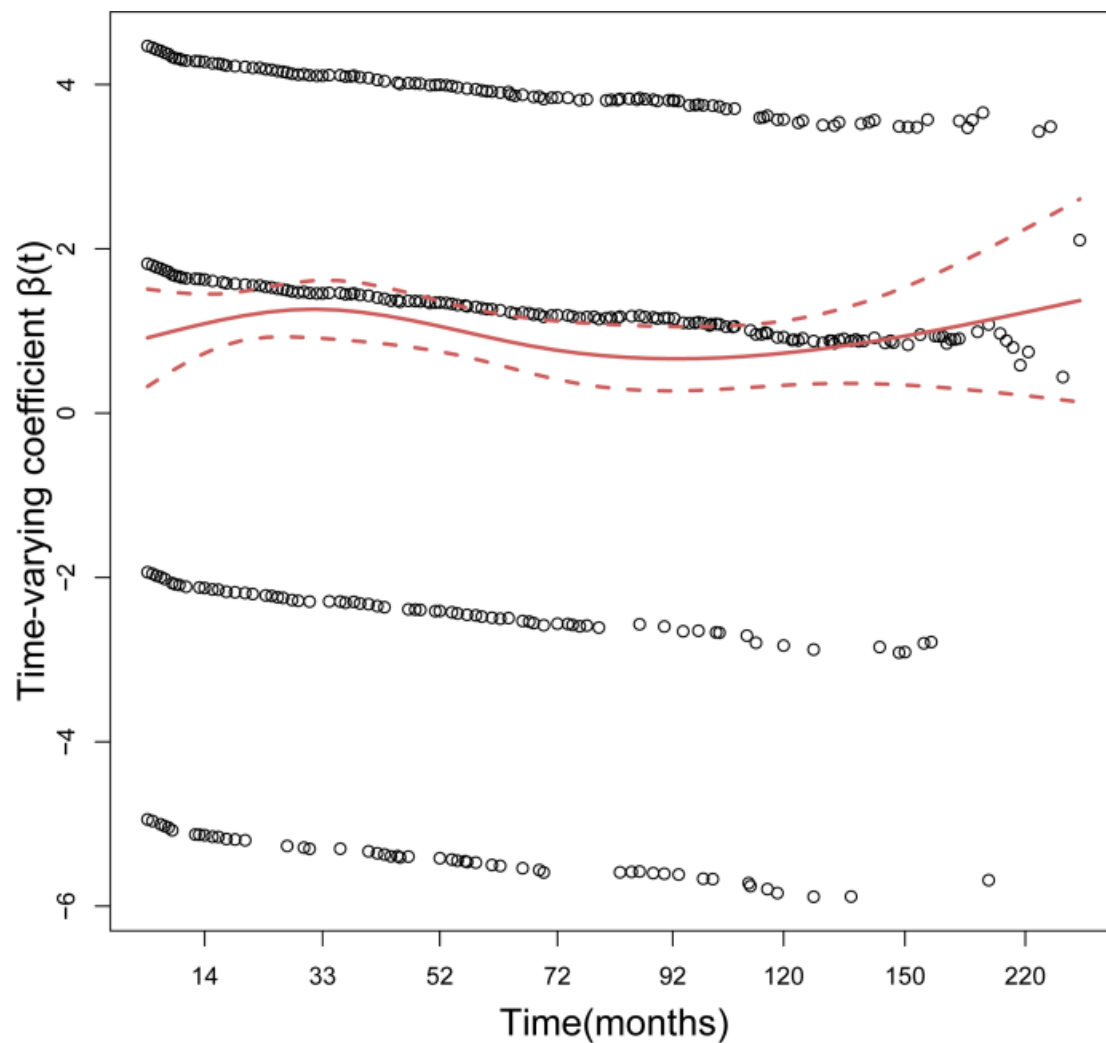


Figure S1 Proportional hazards assessment for the multivariable Cox model (primary model). Time-varying coefficient estimated from Schoenfeld residuals for regional lymph node yield in the multivariable Cox model. The solid line represents the estimated coefficient over time, and dashed lines indicate the 95% confidence bands.

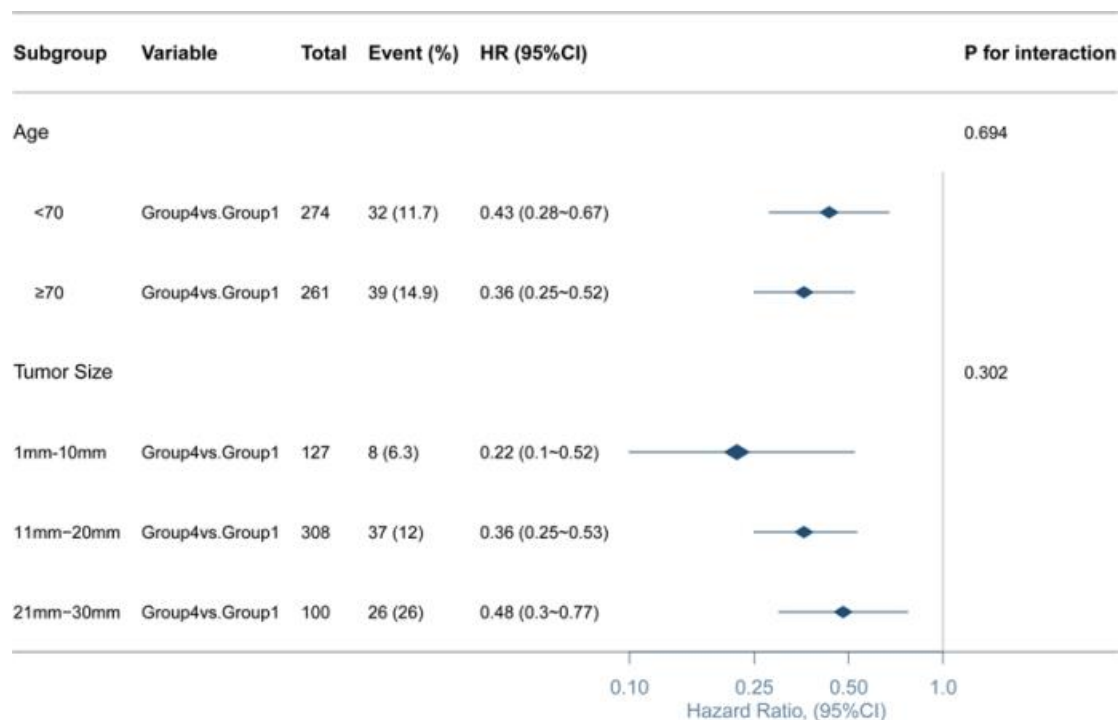


Figure S2 Subgroup analyses of the association between lymph node yield and overall survival. Adjusted hazard ratios (HR) and 95% confidence intervals (CIs) for Group 4 (>10 examined lymph nodes) versus Group 1 (0 nodes) across prespecified subgroups of age and tumor size. Multivariable Cox proportional hazards models were adjusted for age, sex, race, tumor size, tumor grade, histology, laterality, primary lobe, radiotherapy, and chemotherapy. P for interaction was derived from multiplicative interaction terms.