Supplementary Information

Table S1. Data Provenance, Definitions, and Analytical Availability of Variables Initially Considered for the Global Ecological Analysis of Lung Cancer Incidence.

Category	Variable	Source / Access platform	Reference year	Definition / Unit	Available observations, N	Missing observations, n (%)	Presented in S2
Cancer outcomes	Lung cancer incidence—total population	IARC / Cancer Today, GLOBOCAN 2024	2024	Age-standardized incidence rate / cases per 100,000 population	181	2 (1.09%)	Outcome
Cancer outcomes	Lung cancer incidence—males	IARC / Cancer Today, GLOBOCAN 2024	2024	Age-standardized incidence rate / cases per 100,000 males	182	1 (0.55%)	Outcome
Cancer outcomes	Lung cancer incidence—females	IARC / Cancer Today, GLOBOCAN 2024	2024	Age-standardized incidence rate / cases per 100,000 females	182	1 (0.55%)	Outcome
Cancer outcomes	Lung cancer incidence—15–64 years	IARC / Cancer Today, GLOBOCAN 2024	2024	Incidence rate for the selected age interval / cases per 100,000 population	182	1 (0.55%)	Outcome
Lifestyle factors	Current tobacco use prevalence	WHO / OWID	2020	Age-standardized prevalence of current tobacco use among persons aged ≥15 years / %	148	35 (19.13%)	Yes
Environmental factors	Nitrogen dioxide (NO_2) concentration	IHME / GBD platform	2020	Ambient concentration / $\mu g/m^3$	179	4 (2.19%)	Yes
Environmental factors	Sulfur dioxide (SO_2) concentration	IHME / GBD platform	2020	Ambient concentration / $\mu g/m^3$	183	0 (0.00%)	Yes
Occupational exposures	Occupational	IHME / GBD platform	2020	Population exposure estimate /	176	7 (3.83%)	Yes

	exposure to silica			source-defined metric			
Occupational exposures	Occupational exposure to asbestos	IHME / GBD platform	2020	Female population exposure estimate / source-defined metric	173	10 (5.46%)	No
Occupational exposures	Occupational exposure to diesel engine exhaust	IHME / GBD platform	2020	Population exposure estimate / source-defined metric	176	7 (3.83%)	No
Socioeconomic factors	Expected years of schooling	UNDP / OWID	2021	Expected years of schooling / years	170	13 (7.10%)	Yes
Socioeconomic factors	Out-of-pocket health expenditure	WHO / World Bank / OWID	2019	Out-of-pocket health expenditure per capita / source-defined currency unit	171	12 (6.56%)	Yes
Socioeconomic factors	Human Development Index	UNDP	2020	Composite development indicator / index	—	—	No
Water and sanitation	Unsafe water source	WHO / UNICEF	2020	Source-defined water exposure indicator	176	7 (3.83%)	Yes
Water and sanitation	Unsafe sanitation	WHO / UNICEF	2020	Source-defined sanitation exposure indicator	176	7 (3.83%)	Yes
Water and sanitation	No access to handwashing facilities	WHO / UNICEF	2020	Population without handwashing facilities / %	176	7 (3.83%)	No
Dietary factors	Calories from meat consumption	FAO / FAOSTAT	2020	Daily caloric availability from meat / kcal/person/day	172	11 (6.01%)	Yes

Dietary factors	Calories from dairy and eggs	FAO / FAOSTAT	2020	Daily caloric availability from dairy and eggs / kcal/person/ day	173	10 (5.46%)	Yes
Dietary factors	Calories from milk consumpti on	FAO / FAOSTAT	2020	Daily caloric availability from milk / kcal/person/ day	—	—	No
Dietary factors	Calories from fats and oils	FAO / FAOSTAT	2020	Daily caloric availability from fats and oils / kcal/person/ day	171	12 (6.56%)	Yes
Dietary factors	Sugar-derived kilocalories	FAO / FAOSTAT	2020	Daily caloric availability from sugar / kcal/person/ day	172	11 (6.01%)	No
Dietary factors	Vegetable supply	FAO / FAOSTAT	2020	Vegetable supply / g/person/day	170	13 (7.10%)	Yes
Dietary factors	Fruit supply	FAO / FAOSTAT	2020	Fruit supply / g/person/day	169	14 (7.65%)	No
Dietary risk factors	Diet high in red meat	IHME / GBD platform	2020	Dietary exposure estimate / source-defined metric	176	7 (3.83%)	No
Dietary risk factors	Diet low in seafood omega-3 fatty acids	IHME / GBD platform	2020	Dietary exposure estimate / source-defined metric	176	7 (3.83%)	Yes
Dietary risk factors	Diet low in milk	IHME / GBD platform	2020	Dietary exposure estimate / source-defined metric	176	7 (3.83%)	Yes
Metabolic risk factors	High body-mass index	IHME / GBD platform	2020	Metabolic exposure estimate / source-defined metric	176	7 (3.83%)	No

Metabolic risk factors	High fasting plasma glucose	IHME / GBD platform	2020	Metabolic exposure estimate / source-defined metric	176	7 (3.83%)	No
Metabolic risk factors	High systolic blood pressure	IHME / GBD platform	2020	Metabolic exposure estimate / source-defined metric	176	7 (3.83%)	No
Behavioral risk factors	Physical inactivity	IHME / GBD platform	2020	Physical inactivity exposure estimate / source-defined metric	—	—	No
Substance use	Alcohol consumption per person	WHO	2018	Pure alcohol consumption among persons aged ≥15 years / L/person/year	171	12 (6.56%)	No
Health indicators	Breast cancer incidence	IHME / GBD platform	2020	Age-standardized incidence rate / cases per 100,000 population	183	0 (0.00%)	No

Available observations refer to countries with numeric values for each variable within the 183-country study dataset. Missing percentages were calculated as (183 − available observations)/183 × 100. These counts differ from correlation sample sizes, which require available values for both the indicator and the corresponding incidence outcome. No missing values were imputed. Reference years refer to the observations used rather than publication, update, or access dates. Presentation in Supplementary Table S2 does not imply inclusion in the multivariable models. IARC, International Agency for Research on Cancer; GBD, Global Burden of Disease; IHME, Institute for Health Metrics and Evaluation; WHO, World Health Organization; OWID, Our World in Data; UNDP, United Nations Development Programme; UNICEF, United Nations Children's Fund; FAO, Food and Agriculture Organization of the United Nations.

Table S2. Exploratory Spearman Correlations between Lung Cancer Incidence and Selected Country-Level Indicators.

Indicator	Total population	Males	Females	15–64 years
Current tobacco use prevalence	0.616 [0.507, 0.705]; p < 0.001; N = 148	0.677 [0.581, 0.755]; p < 0.001; N = 148	0.492 [0.368, 0.594]; p < 0.001; N = 148	0.660 [0.564, 0.737]; p < 0.001; N = 148
NO_2 concentration	0.479 [0.351, 0.595]; p < 0.001; N = 178	0.468 [0.341, 0.582]; p < 0.001; N = 179	0.462 [0.336, 0.574]; p < 0.001; N = 179	0.460 [0.330, 0.575]; p < 0.001; N = 179
SO_2 concentration	0.417 [0.279, 0.541]; p < 0.001; N = 181	0.427 [0.289, 0.551]; p < 0.001; N = 182	0.379 [0.245, 0.502]; p < 0.001; N = 182	0.404 [0.269, 0.529]; p < 0.001; N = 182
Expected years of schooling	0.725 [0.644, 0.787]; p < 0.001; N = 170	0.674 [0.586, 0.740]; p < 0.001; N = 170	0.728 [0.641, 0.793]; p < 0.001; N = 170	0.664 [0.573, 0.737]; p < 0.001; N = 170
Out-of-pocket health expenditure	0.706 [0.621, 0.769]; p < 0.001; N = 171	0.661 [0.574, 0.729]; p < 0.001; N = 171	0.695 [0.602, 0.766]; p < 0.001; N = 171	0.656 [0.566, 0.727]; p < 0.001; N = 171
Calories from meat	0.699 [0.614, 0.763]; p < 0.001; N = 172	0.644 [0.549, 0.720]; p < 0.001; N = 172	0.732 [0.655, 0.790]; p < 0.001; N = 172	0.640 [0.545, 0.716]; p < 0.001; N = 172
Calories from dairy and eggs	0.649 [0.546, 0.734]; p < 0.001; N = 172	0.623 [0.521, 0.707]; p < 0.001; N = 172	0.608 [0.495, 0.704]; p < 0.001; N = 172	0.609 [0.499, 0.697]; p < 0.001; N = 172
Calories from fats and oils	0.558 [0.448, 0.654]; p < 0.001; N = 170	0.509 [0.393, 0.609]; p < 0.001; N = 170	0.552 [0.437, 0.648]; p < 0.001; N = 170	0.514 [0.395, 0.619]; p < 0.001; N = 170
Vegetable supply	0.497 [0.367, 0.606]; p < 0.001; N = 169	0.519 [0.392, 0.625]; p < 0.001; N = 169	0.393 [0.254, 0.516]; p < 0.001; N = 169	0.501 [0.374, 0.609]; p < 0.001; N = 169
Unsafe water source	−0.786 [−0.833, −0.723]; p < 0.001; N = 175	−0.729 [−0.780, −0.659]; p < 0.001; N = 175	−0.785 [−0.839, −0.711]; p < 0.001; N = 175	−0.729 [−0.784, −0.652]; p < 0.001; N = 175
Unsafe sanitation	−0.739 [−0.800, −0.656]; p < 0.001; N = 175	−0.661 [−0.740, −0.560]; p < 0.001; N = 175	−0.763 [−0.822, −0.684]; p < 0.001; N = 175	−0.666 [−0.743, −0.570]; p < 0.001; N = 175
Diet low in seafood omega-3 fatty acids	−0.571 [−0.669, −0.455]; p < 0.001; N = 175	−0.509 [−0.613, −0.385]; p < 0.001; N = 175	−0.627 [−0.717, −0.515]; p < 0.001; N = 175	−0.501 [−0.602, −0.377]; p < 0.001; N = 175
Diet low in milk	−0.597 [−0.688, −0.490]; p < 0.001; N = 175	−0.583 [−0.669, −0.479]; p < 0.001; N = 175	−0.544 [−0.644, −0.430]; p < 0.001; N = 175	−0.578 [−0.673, −0.468]; p < 0.001; N = 175
Occupational exposure to silica	0.227 [0.085, 0.368]; p = 0.003; N = 175	0.194 [0.045, 0.337]; p = 0.010; N = 175	0.278 [0.144, 0.406]; p < 0.001; N = 175	0.211 [0.066, 0.349]; p = 0.005; N = 175

Values are Spearman's rank correlation coefficients (ϱ), followed by their 95% confidence intervals, two-sided p-values, and analytical sample sizes (N). Spearman correlation coefficients are reported with 95% percentile bootstrap confidence intervals based on 10,000 resamples, two-sided p-values, and the number of available paired observations. P-values below 0.001 are reported as p < 0.001; other values are reported to three decimal places. No adjustment for multiple comparisons was applied across indicators and the four incidence outcomes. Accordingly, p < 0.05 was treated as a nominal threshold, and the analyses were interpreted as exploratory and hypothesis-generating.

Table S3. Countries with the Highest and Lowest Available Values of Selected Country-Level Indicators and Corresponding Lung Cancer Incidence Rates.

Factor	**Exposure Category**	**Country**	**Exposure Value**	**Lung Cancer Incidence**
Tobacco consumption	Highest	Türkiye	67.99	32.02
		Hungary	64.36	40.65
		Serbia	59.60	39.73
		Bosnia and Herzegovina	59.51	33.45
		Montenegro	58.77	33.81
	Lowest	Niger	1.83	0.76
		Nigeria	1.90	1.12
		Benin	2.07	1.05
		Angola	3.28	1.97
		Burkina Faso	3.01	1.68
Expected years of schooling	Highest	Australia	21.05	14.64
		Belgium	19.60	22.84
		Austria	16.01	21.25
		Canada	16.40	17.11
		Argentina	17.87	14.38
	Lowest	Chad	8.04	1.68
		Burkina Faso	9.14	1.68
		Afghanistan	10.26	6.11
		Burundi	10.72	2.50
		Benin	10.77	1.05
Personal health expenditure	Highest	Armenia	1370.43	20.84
		Austria	1169.90	21.25
		Belgium	1062.55	22.84
		Australia	846.11	14.64
		Canada	823.39	17.11
	Lowest	Burundi	15.25	2.50
		Benin	38.57	1.05
		Burkina Faso	42.47	1.68
		Angola	66.70	1.97
		Chad	ND	1.68
Calories from meat consumption	Highest	Argentina	629.22	14.38
		Australia	532.45	14.64
		Belgium	515.24	22.84
		Brazil	486.54	10.12
		Bahamas	481.78	5.75
	Lowest	Burundi	21.41	2.50
		Chad	45.40	1.68
		Bangladesh	65.91	7.08
		Benin	76.71	1.05
		Cameroon	83.17	2.32
Nitrogen dioxide (NO_2) concentration	Highest	Bangladesh	36.86	7.08
		Belarus	32.00	23.11

		Austria	26.58	21.25
		Belgium	23.76	22.84
		Canada	18.67	17.11
	Lowest	Angola	0.00	1.97
		Cameroon	0.00	2.32
		Belize	1.00	6.99
		Bolivia	1.00	4.52
		Brunei Darussalam	1.00	13.56
Occupational exposure to silica	Highest	Cambodia	89.63	13.82
		Bangladesh	89.02	7.08
		Burundi	78.75	2.50
		Benin	73.86	1.05
		Burkina Faso	72.78	1.68
	Lowest	Australia	12.10	14.64
		Belgium	7.67	22.84
		Austria	10.25	21.25
		Canada	15.94	17.11
		Switzerland	ND	16.86
Unsafe water source	Highest	Afghanistan	79.08	6.11
		Burundi	79.29	2.50
		Burkina Faso	78.39	1.68
		Bangladesh	72.74	7.08
		Cameroon	66.48	2.32
	Lowest	Australia	2.20	14.64
		Austria	0.25	21.25
		Belgium	0.33	22.84
		Canada	0.36	17.11
		Switzerland	ND	16.86
Unsafe sanitation	Highest	Burkina Faso	86.63	1.68
		Burundi	82.67	2.50
		Benin	78.79	1.05
		Afghanistan	68.47	6.11
		Cameroon	65.58	2.32
	Lowest	Australia	1.74	14.64
		Belgium	0.88	22.84
		Austria	2.94	21.25
		Canada	1.92	17.11
		Switzerland	ND	16.86
Diet low in seafood omega-3 fatty acids	Highest	Cambodia	89.63	13.82
		Bangladesh	89.02	7.08
		Bolivia	76.14	4.52
		Afghanistan	69.49	6.11
		Bhutan	60.19	4.62
	Lowest	Australia	12.10	14.64
		Austria	10.25	21.25
		Belgium	7.67	22.84
		Canada	15.94	17.11
		Japan	ND	18.70
Diet low in milk	Highest	Cambodia	99.87	13.82

		Benin	99.17	1.05
		Burkina Faso	94.14	1.68
		Bangladesh	93.90	7.08
		Afghanistan	82.70	6.11
	Lowest	Australia	17.54	14.64
		Austria	44.78	21.25
		Belgium	52.10	22.84
		Canada	48.70	17.11
		Netherlands	ND	23.60
Countries with the highest and lowest exposure levels for determinants consistently associated with trachea, bronchus, and lung cancer incidence in the global ecological analysis. Exposure values correspond to the original metrics reported by each data source. Lung cancer incidence is expressed as age-standardized incidence rates per 100,000 population. ND indicates data not available.				

Table S4. Collinearity Diagnostics Across Demographic-Specific Multivariable Regression Models

Variable	Total Population VIF	Male VIF	Female VIF	15–64 Years VIF
Tobacco consumption per person	1.164	1.164	1.164	1.164
Nitrogen dioxide (NO_2) concentration	1.379	1.379	1.379	1.379
Sulfur dioxide (SO_2) concentration	1.295	1.295	1.295	1.295
Calories from meat consumption	2.100	2.100	2.100	2.100
Expected years of schooling	2.287	2.287	2.287	2.287
Personal health expenditure	2.096	2.096	2.096	2.096
Collinearity diagnostics were calculated for the predictors included in the multivariable regression models for the total population, males, females, and individuals aged 15–64 years.				

Table S5. Tolerance Statistics Across Demographic-Specific Multivariable Regression Models

Variable	Total Population	Male	Female	15–64 Years
Tobacco consumption per person	0.859	0.859	0.859	0.859
Nitrogen dioxide (NO_2) concentration	0.725	0.725	0.725	0.725
Sulfur dioxide (SO_2) concentration	0.772	0.772	0.772	0.772
Calories from meat consumption	0.476	0.476	0.476	0.476
Expected years of schooling	0.437	0.437	0.437	0.437
Personal health expenditure	0.477	0.477	0.477	0.477
Tolerance values >0.20 and variance inflation factor (VIF) values <5 were considered indicative of acceptable collinearity. All predictors demonstrated acceptable collinearity diagnostics, indicating that multicollinearity was unlikely to influence model estimates.				

Table S6. Additional GDP-Adjusted Regression Models, Diagnostics, and Sensitivity Analyses for Lung Cancer Incidence.

Panel A. GDP-adjusted regression models						
Outcome	**Predictor**	**B**	**SE HC3**	**95% CI for B**	**Standardized β**	**p-value**
Total	Tobacco use prevalence	0.4520	0.0662	0.3210 to 0.5829	0.371	<0.001
Total	NO_2 concentration	−0.0131	0.0430	−0.0982 to 0.0720	−0.021	0.761
Total	SO_2 concentration	0.0536	0.0602	−0.0654 to 0.1726	0.067	0.375
Total	Calories from meat	0.0165	0.0069	0.0028 to 0.0301	0.227	0.019
Total	Expected years of schooling	0.9229	0.3685	0.1939 to 1.6519	0.231	0.013
Total	Out-of-pocket health expenditure	0.0020	0.0027	−0.0034 to 0.0074	0.064	0.456
Total	ln(GDP per capita)	1.7513	1.1396	−0.5029 to 4.0055	0.173	0.127
Males	Tobacco use prevalence	0.8695	0.1202	0.6317 to 1.1073	0.479	<0.001
Males	NO_2 concentration	−0.0272	0.0667	−0.1591 to 0.1046	−0.029	0.684
Males	SO_2 concentration	0.0792	0.1141	−0.1465 to 0.3048	0.067	0.489
Males	Calories from meat	0.0202	0.0110	−0.0015 to 0.0420	0.187	0.068
Males	Expected years of schooling	1.2287	0.6688	−0.0942 to 2.5515	0.206	0.068
Males	Out-of-pocket health expenditure	0.0011	0.0058	−0.0104 to 0.0125	0.023	0.850
Males	ln(GDP per capita)	2.0248	1.6356	−1.2107 to 5.2602	0.134	0.218
Females	Tobacco use prevalence	0.1224	0.0499	0.0237 to 0.2211	0.133	0.016

Females	NO_2 concentration	−0.0027	0.0377	−0.0772 to 0.0718	−0.006	0.943
Females	SO_2 concentration	0.0249	0.0439	−0.0620 to 0.1117	0.041	0.572
Females	Calories from meat	0.0142	0.0055	0.0033 to 0.0251	0.258	0.011
Females	Expected years of schooling	0.6897	0.2881	0.1197 to 1.2596	0.228	0.018
Females	Out-of-pocket health expenditure	0.0031	0.0022	−0.0013 to 0.0075	0.128	0.167
Females	ln(GDP per capita)	1.3929	1.0527	−0.6894 to 3.4753	0.182	0.188
15–64 years	Tobacco use prevalence	0.4296	0.0664	0.2983 to 0.5610	0.444	<0.001
15–64 years	NO_2 concentration	−0.0134	0.0384	−0.0894 to 0.0626	−0.027	0.728
15–64 years	SO_2 concentration	0.0665	0.0639	−0.0598 to 0.1929	0.105	0.299
15–64 years	Calories from meat	0.0123	0.0062	−0.00002 to 0.02469	0.214	0.0503
15–64 years	Expected years of schooling	0.5812	0.3217	−0.0553 to 1.2176	0.183	0.073
15–64 years	Out-of-pocket health expenditure	0.0007	0.0021	−0.0034 to 0.0049	0.029	0.725
15–64 years	ln(GDP per capita)	0.9664	0.8950	−0.8040 to 2.7369	0.120	0.282

Panel B. Model comparisons and diagnostics

Outcome	N	Six-predictor adjusted R^2[1]	GDP-adjusted R^2	Robust overall model p	Breusch–Pagan p	Maximum VIF	Influential observations, n
Total	140	0.664	0.668	<0.001	0.005	4.690	12
Males	140	0.615	0.616	<0.001	0.005	4.690	11
Females	140	0.558	0.562	<0.001	<0.001	4.690	11
15–64 years	140	0.584	0.584	<0.001	<0.001	4.690	13

Panel C. Sensitivity analyses after excluding influential observations. Each predictor cell reports B (p-value).

Predictor / Model statistic	Total	Males	Females	15–64 years
Tobacco use prevalence	0.3863 (<0.001)	0.8438 (<0.001)	0.1309 (0.003)	0.3663 (<0.001)
NO_2 concentration	−0.0267 (0.355)	−0.0333 (0.591)	0.0131 (0.675)	−0.0235 (0.304)
SO_2 concentration	0.0513 (0.160)	0.0604 (0.529)	0.0046 (0.928)	0.0445 (0.101)
Calories from meat	0.0146 (0.009)	0.0255 (0.003)	0.0142 (0.002)	0.0143 (0.006)
Expected years of schooling	0.6535 (0.033)	0.4811 (0.341)	0.6023 (0.012)	0.4894 (0.069)
Out-of-pocket health expenditure	0.0031 (0.263)	0.0011 (0.784)	0.0033 (0.080)	0.0006 (0.747)
ln(GDP per capita)	2.6737 (0.013)	3.3172 (0.031)	0.7007 (0.389)	1.2226 (0.139)
Remaining N	128	129	129	127
Adjusted R^2	0.729	0.680	0.600	0.663

GDP-adjusted models included the six original predictors and the natural logarithm of GDP per capita at purchasing power parity for 2020, expressed in current international dollars. All primary additional models included 140 countries with complete data. B denotes the unstandardized coefficient; SE HC3 denotes the heteroscedasticity-consistent standard error; β denotes the standardized coefficient; CI denotes confidence interval; VIF denotes variance inflation factor. Panel A confidence intervals and p-values use HC3 robust inference with the t distribution. Panel C uses HC3 robust inference with the asymptotic normal distribution. Influential observations were defined by Cook's distance >4/N and excluded only from sensitivity analyses. P-values are two-sided and unadjusted for multiple comparisons. The p-value of 0.0503 is above the nominal 0.05 threshold. Intercepts were included in all models but are omitted from Panel A. [1] Six-predictor models were refitted to the same 140-country sample; these adjusted R^2 values therefore differ from those of the original full-sample models.

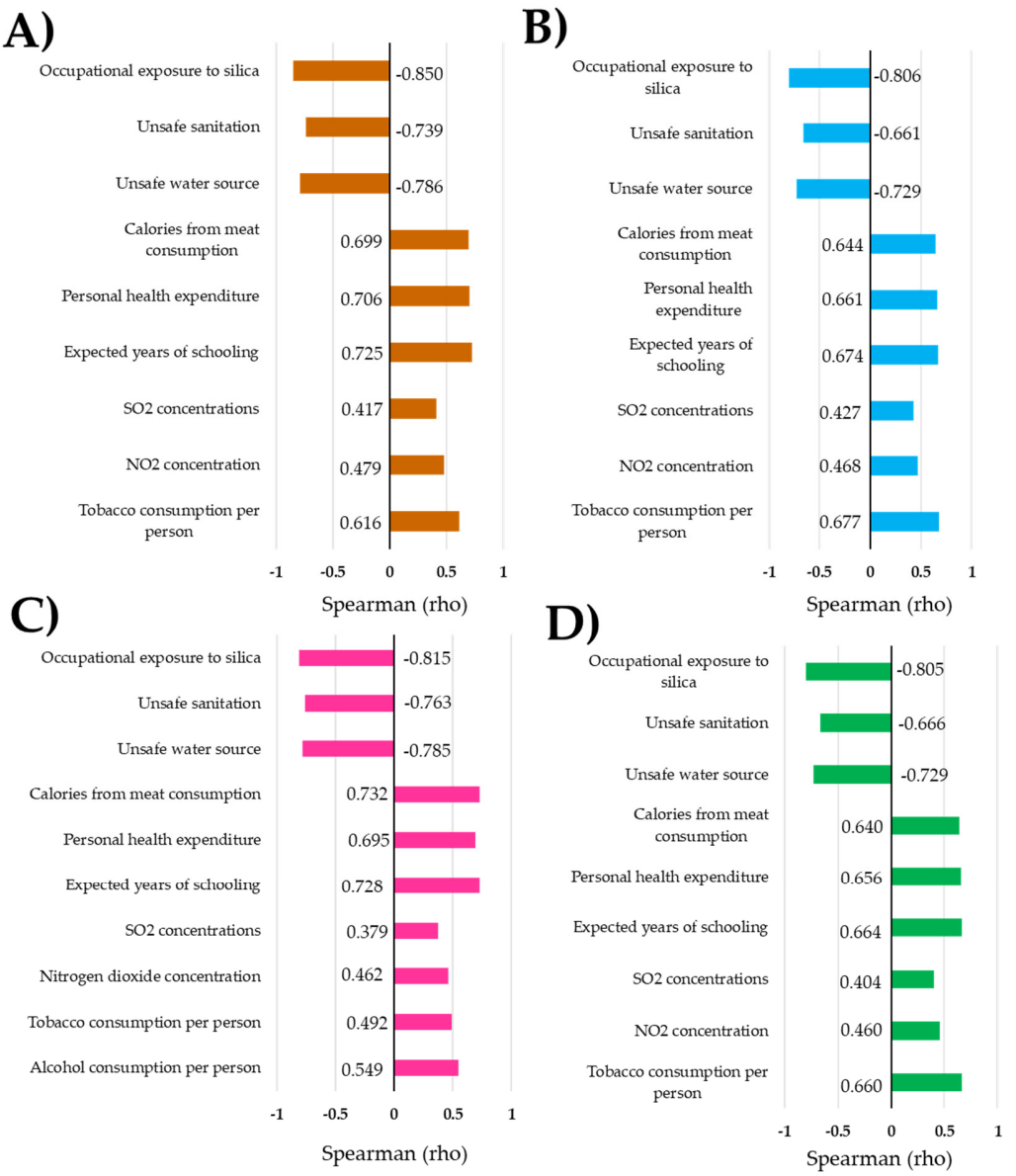


Figure S1. Spearman's rank correlation coefficients (ϱ) between selected socioeconomic, environmental, and lifestyle factors and trachea, bronchus, and lung cancer incidence according to demographic groups. (A) total population incidence, (B) male incidence, (C) female incidence, and (D) incidence among individuals aged 15–64 years. Positive values indicate direct correlations, whereas negative values indicate inverse correlations.

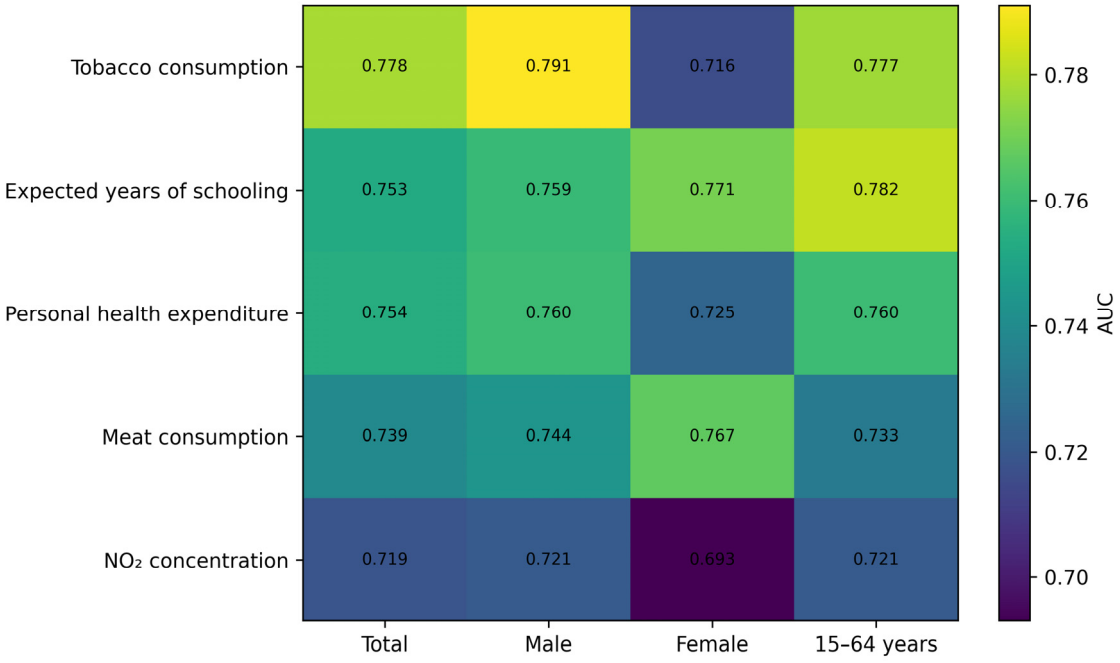


Figure S2. Heatmap illustrating the area under the receiver operating characteristic curve (AUC) for selected socioeconomic, environmental, and dietary determinants associated with trachea, bronchus, and lung cancer incidence across demographic groups. Columns represent the total population, males, females, and individuals aged 15–64 years. Color intensity reflects discriminatory performance, with higher AUC values indicating greater ability to distinguish countries with high versus low lung cancer incidence. Tobacco consumption, expected years of schooling, personal health expenditure, and meat consumption consistently demonstrated the highest discriminatory capacity across all demographic groups, whereas nitrogen dioxide (NO_2) concentration showed comparatively lower but still acceptable discriminatory performance. All AUC values were significantly different from 0.50 ($p < 0.001$).