

Smoking Prevalence and Lung Cancer Incidence Across Texas Counties

A pre-specified county-level comparison of CDC PLACES adult smoking estimates with Texas Cancer Registry lung cancer incidence.

Author

Vellora Research

Published

2026-10-23

Type

Ecological, cross-sectional secondary-data analysis; not peer reviewed

Data years

Smoking: BRFSS 2023 (PLACES 2025); lung cancer: diagnoses 2019–2023

Online report, with the interactive figure, data and code: vellora-health.pages.dev/research/reports/smoking-lung-cancer-texas/

Abstract

Background. Cigarette smoking is the leading risk factor for lung cancer. ^[1, 2] This study asks whether Texas counties with higher estimated adult smoking prevalence also have higher lung cancer incidence.

Methods. Ecological, cross-sectional analysis of CDC PLACES 2025 age-adjusted current smoking prevalence among adults (BRFSS 2023) and State Cancer Profiles age-adjusted lung and bronchus cancer incidence (Texas Cancer Registry, 2019–2023). The analysis was specified before results were computed: one confirmatory test, Spearman’s rank correlation with a 95% confidence interval, and five labelled exploratory analyses. Of 254 counties, 203 had both values published; 51 were excluded because the lung cancer rate (and, for one county, the smoking estimate) is suppressed by the source.

Results. Median county smoking prevalence was 15.4% (interquartile range 13.8–16.6) and median lung cancer incidence 52.4 per 100,000 (IQR 43.0–63.5). Smoking and lung cancer incidence were positively associated ($\rho = 0.44$, 95% CI 0.32 to 0.55, $p < 0.0001$, $n = 203$), a moderate correlation. In exploratory analyses the association was stronger among metropolitan ($\rho = 0.68$, 95% CI 0.52 to 0.79) than nonmetropolitan counties ($\rho = 0.32$, 95% CI 0.14 to 0.47) and changed little when any single county was removed (ρ 0.43 to 0.46).

Conclusions. Across Texas counties with published data, higher estimated adult smoking prevalence went with higher lung cancer incidence. The finding is consistent with the established individual-level evidence on smoking and lung cancer, but this county-level, cross-sectional design cannot show that smoking caused any county's rate, and it does not apply to individuals.

1. Introduction

CDC describes cigarette smoking as the number one risk factor for lung cancer, linked to about 80% of U.S. lung cancer deaths; radon, secondhand smoke, workplace exposures such as asbestos and diesel exhaust, and air pollution are other risk factors. [1] The causal role of smoking in lung cancer is established by decades of individual-level evidence summarised by the U.S. Surgeon General. [2]

This report does not test that causal relationship, which is not in question. It asks a narrower, descriptive question about Texas: do the county-level figures published by CDC and the Texas Cancer Registry show the pattern one would expect, and how strong is it once the uncertainty, suppression and design limits of county data are taken into account?

Research question: Is age-adjusted adult current cigarette smoking prevalence associated with age-adjusted lung and bronchus cancer incidence across Texas counties?

2. Data and methods

Table 1. Measures analysed

Measure	Definition	Years and adjustment	Source
Current cigarette smoking among adults (exposure)	Adults 18 and over who have smoked at least 100 cigarettes in their lifetime and now smoke every day or some days, % (model-based small-area estimate)	BRFSS 2023; age-adjusted prevalence (data value type AgeAdjPrv; CDC does not publish the standard population used); 95% CI published	CDC PLACES, County Data, 2025 release [3] , 4 , 6]
Lung and bronchus cancer incidence (outcome)	New invasive lung and bronchus cancers, all stages, all ages, both sexes, per 100,000 people per year (registry counts)	2019–2023 average annual rate; age-adjusted to the 2000 U.S. standard population; 95% CI published	State Cancer Profiles (NCI and CDC), Texas Cancer Registry data [7] , 8]
Metropolitan status (exploratory grouping)	Rural-Urban Continuum Codes 1–3 metropolitan, 4–9 nonmetropolitan	2023 codes; 2020 Census population	USDA Economic Research Service [9]

How each measure is age-adjusted. Age adjustment recalculates a figure as if every county had the same age mix (a “standard population”), so counties with older or younger residents can be compared fairly. The two measures are adjusted separately and describe different things:

- **Lung cancer incidence** counts new diagnoses among people of all ages, per 100,000 people per year over 2019–2023, and is age-adjusted to the 2000 U.S. standard population. [\[7\]](#), [8\]](#)
- **Smoking prevalence** is the estimated percentage of adults aged 18 and over who currently smoke, in 2023. CDC estimates it with a statistical model (multilevel regression and poststratification) that combines BRFSS 2023 survey answers, including each respondent’s age, sex, race and ethnicity and education, with Census 2023 county

population estimates and American Community Survey 2019–2023 data. [4, 6, 14] CDC publishes county estimates both crude and age-adjusted, and says the adjustment uses a set of age groups. [5, 6]

- **What is not documented.** None of the CDC PLACES documentation reviewed (methodology, measure definitions, data FAQs, 2025 release notes and the dataset description) names the standard population or lists the age groups used for the age-adjusted estimates. [3, 4, 5, 6, 14] Many CDC survey estimates are age-adjusted to the 2000 U.S. standard population, [15] but this could not be confirmed for PLACES from CDC’s own documentation, so this report does not assume it.

Even with the same standard, the two figures would not be on one scale: one is a share of adults in a single year, the other a rate of new cancers among people of all ages over five years. The analysis therefore compares only how counties rank on each measure.

Matching and sample. Counties were matched on their five-digit FIPS codes, which are identical and in the same order in all three verified input files. A county was included only if both values were published. No value was imputed, estimated or replaced with zero; every excluded county is listed with the source’s reason (Table 2).

Pre-specified analysis. Before any association was computed, the analysis code recorded one confirmatory hypothesis (H1: smoking prevalence and lung cancer incidence are positively associated across counties) and its test: Spearman’s rank correlation, [10] chosen because it does not assume a linear relationship or normally distributed values, with a two-sided p-value from the t approximation and a 95% confidence interval from the Fisher z transform with the Bonett–Wright standard error. [11] With a single confirmatory test, no multiple-testing correction is needed ($\alpha = 0.05$).

Exploratory analyses (reported with unadjusted p-values and not used for the conclusion): E1, Pearson’s r with a Fisher z interval; E2, Spearman’s ρ separately in metropolitan and nonmetropolitan counties; E3, the range of ρ when each county is removed in turn (influence); E4, crude rather than age-adjusted smoking prevalence; E5, how excluded counties differ from included ones. No county was removed as an outlier. Counties are unweighted: each counts once regardless of population.

3. Results

3.1 Sample

203 of 254 counties (78 metropolitan, 125 nonmetropolitan) had both values. All 51 exclusions are small counties whose lung cancer rate is suppressed because fewer than 16 cases were diagnosed over 2019–2023; the largest had 7,694 residents in 2020, and the median 2,547, against 24,435 for included counties. 8 of the excluded counties are metropolitan by RUCC. Their median smoking estimate (15.1%) was close to that of included counties (15.4%) (exploratory E5).

Table 2. Counties excluded (51) and the source’s reason

Reason	Counties
Lung cancer: Suppressed by State Cancer Profiles to protect confidentiality and rate stability: fewer than 16 cases were reported over 2019–2023 (average annual count 3 or fewer).	Armstrong, Bailey, Borden, Briscoe, Cochran, Collingsworth, Cottle, Crockett, Culberson, Dallam, Dickens, Edwards, Fisher, Floyd, Foard, Garza, Glasscock, Hall, Hartley, Hemphill, Hudspeth, Irion, Jeff Davis, Jim Hogg, Kenedy, Kent, Kimble, King, Kinney, La Salle, Lipscomb, Lynn, McMullen, Martin, Mason, Menard, Motley, Oldham, Presidio, Reagan, Roberts, Schleicher, Sherman, Sterling, Stonewall, Sutton, Terrell, Throckmorton, Upton, Yoakum (50)
Smoking: Suppressed by CDC: Estimates suppressed for population less than 50. Lung cancer: Suppressed by State Cancer Profiles to protect confidentiality and rate stability: fewer than 16 cases were reported over 2019–2023 (average annual count 3 or fewer).	Loving (1)

Table 3. Descriptive statistics for the 203 included counties (unweighted)

Measure	Mean (SD)	Median (IQR)	Lowest	Highest
Current smoking, adults (%, age-adjusted, 2023)	15.2 (2.2)	15.4 (13.8– 16.6)	9.0 (Collin)	22.9 (Newton, San Augustine)
Lung cancer incidence (per 100,000, age-adjusted, 2019– 2023)	54.0 (15.0)	52.4 (43.0– 63.5)	23.9 (Webb)	89.9 (Polk)

3.2 Primary analysis

County smoking prevalence and lung cancer incidence were positively associated: Spearman $\rho = 0.44$, 95% CI 0.32 to 0.55, two-sided $p < 0.0001$, $n = 203$ (Figure 1). By common rules of thumb this is a moderate correlation: counties ranked higher on smoking tended to rank higher on lung cancer incidence, with considerable scatter. H1 is supported.

Adult smoking and lung cancer incidence by Texas county

Spearman $\rho = 0.44$ (95% CI 0.32 to 0.55), $n = 203$ counties; 51 counties excluded because a value is suppressed

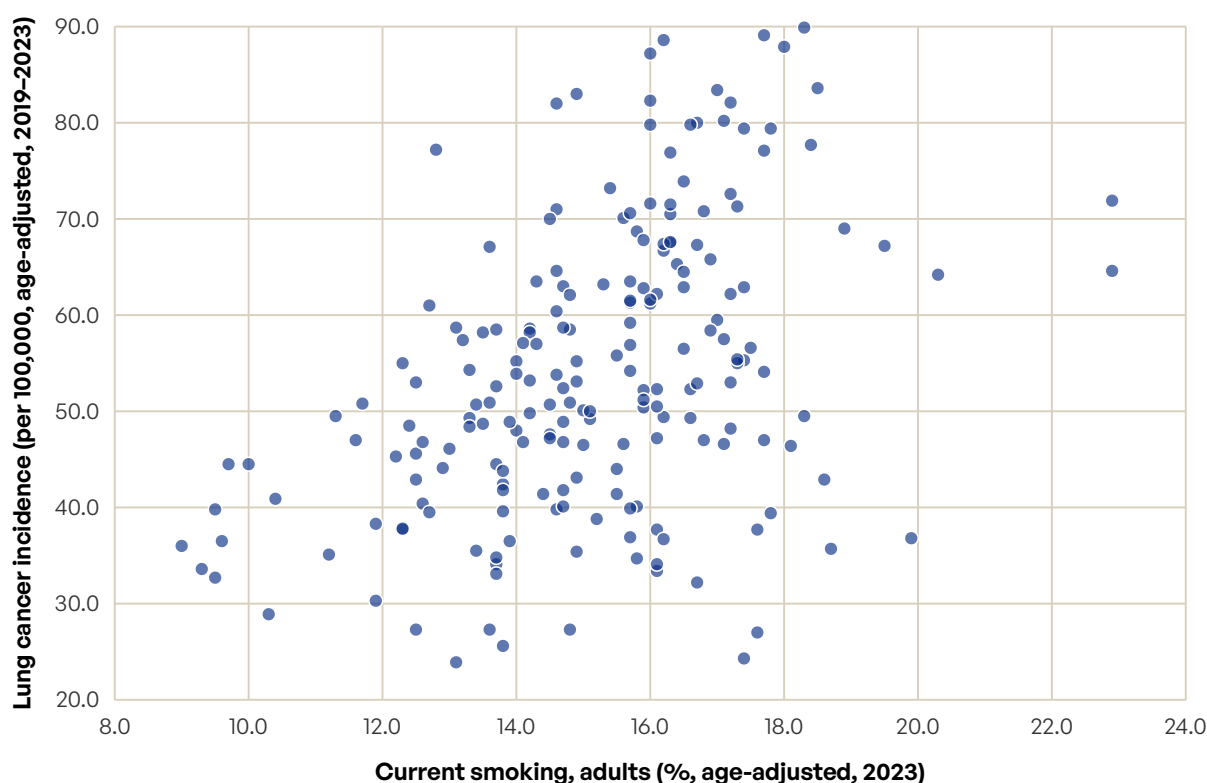


Figure 1. Each point is one of the 203 counties with both values published; in the online report, hover or tap a point, or choose a county, to identify it. Smoking is a CDC PLACES model estimate for 2023; lung cancer incidence is a 2019–2023 average from the Texas Cancer Registry. The 51 counties whose lung cancer rate is suppressed (fewer than 16 cases) are not shown and are listed in Table 2. The pattern is an association between counties, not evidence that smoking caused any county’s rate.

3.3 Exploratory analyses

The Pearson correlation was similar to the Spearman correlation ($r = 0.44$). The association was stronger among the 78 metropolitan counties than among the 125 nonmetropolitan counties; this difference was not formally tested, and nonmetropolitan counties have smaller populations and wider confidence intervals. Removing any one county changed ρ only to between 0.43 (without Brewster) and 0.46 (without Starr), so no single county drives the result. Using crude instead of age-adjusted smoking prevalence gave $\rho = 0.45$.

Table 4. Correlation between smoking prevalence and lung cancer incidence

Analysis	Statistic	Counties	n	Estimate (95% CI)	p (unadjusted)
Primary (H1)	Spearman ρ	All included counties, age-adjusted smoking	203	0.44 (0.32 to 0.55)	< 0.0001
Exploratory E1	Pearson r	All included counties, age-adjusted smoking	203	0.44 (0.32 to 0.54)	< 0.0001
Exploratory E2	Spearman ρ	Metropolitan counties only	78	0.68 (0.52 to 0.79)	< 0.0001
Exploratory E2	Spearman ρ	Nonmetropolitan counties only	125	0.32 (0.14 to 0.47)	0.0003
Exploratory E4	Spearman ρ	Crude (not age- adjusted) smoking	203	0.45 (0.32 to 0.56)	< 0.0001

Exploratory E3 (leave-one-out ρ range 0.43 to 0.46) and E5 (excluded counties) are described in the text.

4. Discussion

Texas counties with higher estimated adult smoking prevalence tended to have higher lung cancer incidence. The direction agrees with the individual-level evidence, [\[1, 2\]](#) and the association held in each exploratory check. Its moderate size is not surprising for county data: smoking is measured in 2023 while lung cancer reflects exposures over earlier decades, smoking estimates are modelled rather than counted, and other causes of lung cancer, such as radon and workplace exposures, [\[1\]](#) vary between counties.

The result should not be read as an estimate of how much of any county's lung cancer is caused by smoking, or of the effect of reducing smoking. An ecological correlation can differ in size and even direction from the individual-level relationship, [\[12, 13\]](#) and county smoking prevalence is correlated with many other county characteristics. The stronger association among metropolitan counties is exploratory and may reflect their larger populations and more precise estimates rather than any real difference.

5. Limitations

- **Ecological fallacy.** The unit is the county. A county-level association says nothing about whether the people who smoke in a county are the people diagnosed with lung cancer there, and cannot be applied to individuals. ^[12, 13]
- **No individual linkage.** Smoking comes from a survey-based model and lung cancer from a cancer registry. No person's smoking history is linked to their diagnosis.
- **Confounding.** Counties differ in age structure beyond what age adjustment removes, income, education, occupation, radon, air pollution, access to screening and diagnosis, and other factors that may relate to both measures. ^[1] None is adjusted for here.
- **Different periods and latency.** Smoking is estimated for 2023 and lung cancer averaged over 2019–2023. Lung cancer usually develops after many years of smoking, so current prevalence stands in for past exposure, which it may not track well where smoking has changed. ^[2] This is a cross-sectional comparison, not a longitudinal study.
- **Model-based smoking estimates.** PLACES county values are small-area model estimates built from national survey data and county demographics, not direct measurements, and carry wide confidence intervals in small counties. ^[4] Part of the between-county variation they show comes from the model's demographic inputs.
- **Suppression and selection.** 51 small counties with fewer than 16 cases are excluded, so the results describe the 203 counties with published rates, which are larger on average. The associations might differ among the smallest counties.
- **Unequal populations and uncertainty.** Each county counts once, from counties of a few thousand residents to Harris County's 4.7 million, so results describe counties, not the average Texan. County estimates are not equally precise, and the correlation does not weight them by precision.
- **Geographic aggregation.** County boundaries are administrative; a different geographic unit could give a different correlation. Neighbouring counties are not independent, which can make p-values look more certain than they are.
- **Different age adjustments.** Smoking is adjusted among adults with a standard population CDC does not document; cancer incidence across all ages to the 2000 U.S. standard. The measures are compared only by rank.

- **Not peer reviewed.** This is an independent analysis, checked by an independent recomputation but not by external reviewers.

6. Conclusion

Among the 203 Texas counties with published data, higher estimated adult smoking prevalence was associated with higher lung cancer incidence ($p = 0.44$, 95% CI 0.32 to 0.55). The pattern matches what the established evidence on smoking would predict, but this exploratory, ecological, cross-sectional analysis does not show that smoking caused any county's lung cancer rate, does not estimate the effect of changing smoking, and does not describe any individual's risk. It is not medical advice.

Independent exploratory analysis — not peer reviewed. This report is by Vellora Research, is not affiliated with any university or public agency, and is not medical advice. CDC, NCI, the Texas Cancer Registry and USDA data are used without endorsement by any of them.

7. References

- [1] Centers for Disease Control and Prevention. Lung Cancer Risk Factors. Accessed 2026-10-23. <https://www.cdc.gov/lung-cancer/risk-factors/index.html>
- [2] U.S. Department of Health and Human Services. The Health Consequences of Smoking —50 Years of Progress: A Report of the Surgeon General. Atlanta, GA: Centers for Disease Control and Prevention; 2014. <https://www.ncbi.nlm.nih.gov/books/NBK179276/>
- [3] Centers for Disease Control and Prevention. PLACES: Local Data for Better Health, County Data, 2025 release (dataset swc5-untb; rows updated 2025-12-04). Accessed 2026-10-22. <https://data.cdc.gov/d/swc5-untb>
- [4] Centers for Disease Control and Prevention. PLACES: Methodology. Accessed 2026-10-23. <https://www.cdc.gov/places/methodology/index.html>
- [5] Centers for Disease Control and Prevention. PLACES Data FAQs: “Why are age-adjusted estimates not provided for all levels?” Last reviewed 2024-08-29. Accessed 2026-10-23. <https://www.cdc.gov/places/faqs/data-faqs/index.html>
- [6] Centers for Disease Control and Prevention. PLACES Measure Definitions: Health Risk Behaviors. Accessed 2026-10-22. <https://www.cdc.gov/places/measure-definitions/health>

[-risk-behaviors.html](#)

[7] National Cancer Institute and Centers for Disease Control and Prevention. State Cancer Profiles: Incidence Rate Report for Texas by County, Lung & Bronchus (All Stages), 2019–2023. Data: Texas Cancer Registry, SEER November 2025 submission. Accessed 2026-10-18. <https://statecancerprofiles.cancer.gov/incidencerates/index.php?stateFIPS=48&areatype=county&cancer=047&race=00&sex=0&age=001&stage=999&year=0&type=incd&sortVariableName=rate&sortOrder=default>

[8] National Cancer Institute, Surveillance, Epidemiology, and End Results Program. Standard Populations (Millions) for Age-Adjustment. Accessed 2026-10-23. <https://seer.cancer.gov/stdpopulations/>

[9] U.S. Department of Agriculture, Economic Research Service. Rural-Urban Continuum Codes, 2023 (data and documentation). Accessed 2026-10-22. <https://www.ers.usda.gov/data-products/rural-urban-continuum-codes>

[10] Spearman C. The proof and measurement of association between two things. Am J Psychol. 1904;15(1):72–101. <https://doi.org/10.2307/1412159>

[11] Bonett DG, Wright TA. Sample size requirements for estimating Pearson, Kendall and Spearman correlations. Psychometrika. 2000;65(1):23–28. <https://doi.org/10.1007/BF02294183>

[12] Robinson WS. Ecological correlations and the behavior of individuals. Am Sociol Rev. 1950;15(3):351–357. <https://doi.org/10.2307/2087176>

[13] Morgenstern H. Ecologic studies in epidemiology: concepts, principles, and methods. Annu Rev Public Health. 1995;16:61–81. <https://doi.org/10.1146/annurev.pu.16.050195.000425>

[14] Centers for Disease Control and Prevention. PLACES: Current Release Notes (2025 release). Accessed 2026-10-23. <https://www.cdc.gov/places/current-release-notes/index.html>

[15] National Center for Health Statistics. Age adjustment. Health, United States: Sources and Definitions. Accessed 2026-10-23. <https://www.cdc.gov/nchs/hus/sources-definition/s/age-adjustment.htm>

8. Reproducibility

Every number in this report, its PDF and Figure 1 is read from one results file, generated by the pre-specified analysis code from Vellora’s verified copies of the source files. The analysis is deterministic. An independent Python implementation using SciPy and NumPy recomputed the sample, every descriptive statistic, correlation, confidence interval, p-value and leave-one-out value and matched the results file.

Results (JSON)

Every statistic reported here, the excluded counties with their reasons, and the per-county values in Figure 1

Analysis code (TypeScript)

The pre-specified analysis that produced the results file

Independent check (Python)

Recomputes every statistic with SciPy and NumPy

Smoking input (JSON)

Vellora's verified copy of CDC PLACES 2025 age-adjusted and crude estimates (indicators smoking-aa and smoking)

Lung cancer input (JSON)

Vellora's verified copy of State Cancer Profiles county incidence (indicator cancer-inc-lung)

County classification input (JSON)

USDA ERS RUCC 2023 codes and 2020 Census populations

To reproduce: regenerate the results with `npm run research:lung` and check them with `uvx --with scipy --with numpy python scripts/verify-lung-study.py`. The input files record the source releases and download dates; CDC and NCI may revise their data.